Viral Etiology of Central Nervous System Infections and Community-Acquired Sepsis in Southeast Asia: Unravelling the Unknown

Thesis

VIRAL ETIOLOGY OF CENTRAL NERVOUS SYSTEM INFECTIONS AND COMMUNITY-ACQUIRED SEPSIS IN SOUTHEAST ASIA: UNRAVELLING THE UNKNOWN

by

NGUYEN TO ANH

A thesis submitted to the Open University, U.K

For the degree of Doctor of Philosophy in the field of Life Science

Oxford University Clinical Research Unit

Hospital for Tropical Diseases

Ho Chi Minh City, Viet Nam

Jan 2021

Abstract

My PhD research consists of a series of studies on the application of metagenomic next-generation sequencing (mNGS) to search for viral etiology in patients presenting with community-acquired sepsis or central nervous system (CNS) infections. I first developed a mNGS workflow for the sensitive detection of a broad range of viruses in clinical samples (Chapter 2). I then used this optimized method to search for viruses in 665 patients presenting with community-acquired sepsis of unknown origin enrolled in an observational study across Thailand and Vietnam in 2013-2015. While the mNGS analysis revealed significant insights into the epidemiology of sepsis in both countries, the analysis also led to the first detection of a recently discovered flavivirus - human pegivirus 2 (HPgV-2) - in a serum sample of a Vietnamese patient co-infected with HIV and HCV. This represents the first detection of HPgV-2 in Vietnam. Therefore, I conducted further research to unravel its epidemiology in Vietnam (Chapter 4). In Chapter 5, I used mNGS to analyze 204 cerebrospinal fluid (CSF) from patients with CNS infections of unknown origin enrolled from hospitals across central and southern Vietnam in 2012-2016. Enteroviruses were the most common viruses detected, especially in children and young adults. To inform future research directions, I conducted a pilot of 66 consecutive CSF samples collected from patients presenting with CNS infections admitted to the Hospital for Tropical Diseases in Ho Chi Minh City, Vietnam (Chapter 6). mNGS accurately detected a wide range of pathogens that were also detected by routine diagnostic methods, but also increased the diagnostic yield from 22.7% (15/66) to 34.8% (23/66) (Chapter 6). Finally, in Chapter 7, I provide an overview about my research findings, and propose some future directions based on the main findings obtained during my PhD research.

Co-Authorship

The work presented in this thesis is my own and were mostly conducted by me under supervisions of Dr Le Van Tan and Prof Guy Edward Thwaites. Bioinformatic pipeline was carried out in collaboration with Prof Eric Delwart at Blood Systems Research Institute, San Francisco, California, United State, with support of Dr Xutao Deng. Next generation sequencing and PCR confirmation testing of Thailand samples from patients with CA sepsis were conducted in the molecular diagnosis lab of Dr Direk Limmathurotsakul at Mahidol-Oxford Tropical Medical Research Unit, Bangkok, Thailand, with support of Dr Janjira Thaipadungpani.

Acknowledgements

I would like to thank the following people, without whom I would have not been able to complete my project. First of all, I would like to express my sincere gratitude to my supervisors: Dr Le Van Tan and Prof Guy Edward Thwaites for their continuous support and dedicated guidance, patience and ongoing encouragement. They have always been enthusiastically willing to give me invaluable advice throughout my research and during my time of writing this thesis.

I would like to send my special thankfulness to Prof Eric L. Delwart at the Blood Systems Research Institute, CA for his enthusiastic support and immense knowledge. Thanks for having me in your group back in 2017, and for offering me your valuable inputs into my research and the papers we published.

I especially wish to thank Ms Le Nguyen Truc Nhu, Ms Nguyen Thi Thu Hong, Ms Nguyen Thi Han Ny, Ms Le Kim Thanh, Dr Tran Tan Thanh and Assoc. Prof. Catherine Louise Thwaites from Emerging Infections group, OUCRU, and Dr Catherine Anscombe from the Malawi-Liverpool Wellcome Trust Clinical Research Programme for their tremendous help and support whenever I need. Also, I greatly appreciate Dr Direk Limmathurotsakul, Dr Janjira Thaipadungpani and staff from the Mahidol-Oxford Tropical Medical Research Unit for supporting me during my laboratory work there in Thailand. Thanks also must go to Dr Xutao Deng and Mr David Ferrari from the Blood Systems Research Institute, CA for helping me with the primary bioinformatics analysis.

I am particularly grateful to all members of the Southeast Asia Infectious Disease Clinical Research Network and the VIZIONS Consortium for allowing me to use the samples collected as part of these respective projects for my research.

I would like to take this opportunity to say thank you to Dr Leigh Jones and Ms Le Thi Kim Yen from OUCRU Training Committee for their valuable support.

Finally, I would like to give thanks with all my hearts to my husband and my mom for supporting me, and being by my side throughout the tough times.

Publications

A. Publications as parts of my PhD research:

1. **Anh Nguyen To**, Hong Nguyen Thi Thu, Nhu Le Nguyen Truc, Thanh Tran Tan, Lau Chuen-Yen, Limmathurotsakul Direk, Deng Xutao, Rahman Motiur, Chau Nguyen Van Vinh, van Doorn H Rogier, Thwaites Guy, Delwart Eric, and Tan Le Van, 2019. Viruses in Vietnamese Patients Presenting with Community-Acquired Sepsis of Unknown Cause. J. Clin. Microbiol. 57, e00386-19.
2. **Anh Nguyen To**, Nhu Le N, Hong Nguyen, Phuc Tran, Tam Pham, Huong Dang Thao, Anh Tran Tuan, Deng Xutao, Nghia Ho D, Nguyen Tran Thua, Van Hung Nguyen, Thuan Nguyen Dac, Phuong Pham T, Chau Nguyen, Baker Stephen, Delwart Eric, Thwaites Guy, and Van Tan Le, 2021. Viral Metagenomic Analysis of Cerebrospinal Fluid from Patients with Acute Central Nervous System Infections of Unknown Origin, Vietnam. Emerg. Infect. Dis. J. 27.
3. Hong Nguyen Thi Thu, **Anh Nguyen To**, Mai Nguyen Thi Hoang, Nghia Ho Dang Trung, Nhu Le Nguyen Truc, Thanh Tran Tan, Phu Nguyen Hoan, Deng Xutao, van Doorn H Rogier, van Vinh Chau Nguyen, Delwart Eric, Thwaites Guy, and van Tan Le, 2020. Performance of metagenomic next-generation sequencing for the diagnosis of viral meningoencephalitis in a resource-limited setting. Open Forum Infect. Dis. 7, 1–10.
4. **Nguyen To Anh**, Nguyen Thi Thu Hong, Le Nguyen Truc Nhu, Tran tan Thanh, Anscombe Catherine, Chau Le Ngoc, Tran Thi Thanh thanh, Lau Chuen Yen, Limmathurotsakul Direk, Nguyen Van Vinh Chau, van Doorn H Rogier, Deng Xutao, Rahman Motiur, Delwart Eric, Le Thuy, Thwaites Guy, and Le Van Tan, 2018. Detection and Characterization of Human Pegivirus 2, Vietnam. Emerg. Infect. Dis. 24, 2063–2067.

B. Other Publications:

5. **Nguyen Anh To**, Tran Thanh Tan, Hoang Van Minh Tu, Nghiem Ngoc My, Le Nhu Nguyen Truc, Le Thanh Thi My, Phan Qui Tu, Truong Khanh Huu, Le Nhan Nguyen Thanh, Ho Viet Lu, Do Viet Chau, Ha Tuan Manh, Nguyen Hung Thanh, Nguyen Chau Van Vinh, Thwaites Guy, van Doorn H Rogier, and Le Tan Van, 2016. Development and evaluation of a non-ribosomal random PCR and next-generation sequencing based assay for detection and sequencing of hand, foot and mouth disease pathogens. Virol. J. 13, 125.
6. **Nguyen To Anh**, Le Nguyen Truc Nhu, Hoang Minh Tu Van, Nguyen Thi Thu Hong, Tran Tan Thanh, Vu Thi Ty Hang, Nguyen Thi Han Ny, Lam Anh Nguyet, Tran Thi lan Phuong, Le Nguyen Thanh Nhan, Nguyen Thanh Hung, Truong Huu Khanh, Ha Manh Tuan, Ho Lu Viet, Nguyen Tran Nam, Do Chau Viet, Phan Tu Qui, Wills Bridget, Sabanathan Sarawathy, Nguyen Van Vinh Chau, Thwaites Louise, Doorn H Rogier Van, Thwaites Guy, Rabaa Maia A, and Le Van Tan, 2018. Emerging Coxsackievirus A6 causing Hand, Foot and Mouth Disease, Vietnam. Emerg. Infect. Dis. 24, 17–19.

7. Mcbride Angela, Chau Tran Thi Hong, Hong Nguyen Thi Thu, Mai Nguyen Thi Hoang, **Anh Nguyen To**, Thanh Tran Tan, Van Tran Thi Hue, Xuan Le Thi, Sieu Tran Phu Manh, Thai Le Hong, Chuong Ly Van, Sinh Dinh Xuan, Phong Nguyen Duy, Phu Nguyen Hoan, Day Jeremy, Nghia Ho Dang Trung, Hien Tran Tinh, Chau Nguyen Van Vinh, Thwaites Guy, and Tan Le Van, 2017. Angiostrongylus cantonensis is an important cause of eosinophilic meningitis in Southern Vietnam. Clin. Infect. Dis. 64, 1784–1787.
8. Ny Nguyen Thi Han, **Anh Nguyen To**, Hang Vu Thi Ty, Nguyet Lam Anh, Thanh Tran Tan, Ha Do Quang, Minh Ngo Ngoc Quang, Ha Do Lien Anh, McBride Angela, Tuan Ha Manh, Baker Stephen, Tam Pham Thi Thanh, Phuc Tran My, Huong Dang Thao, Loi Tran Quoc, Vu Nguyen Tran Anh, Hung Nguyen Van, Minh Tran Thi Thuy, Xang Nguyen Van, Dong Nguyen, Nghia Ho Dang Trung, Chau Nguyen Van Vinh, Thwaites Guy, van Doorn H Rogier, Anscombe Catherine, and Le Van Tan, 2017. Enterovirus D68 in Viet Nam (2009-2015). Wellcome Open Res. 2, 41.
9. Thanh Tran Tan, Van Hoang Minh Tu, Hong Nguyen Thi Thu, Nhu Le Nguyen Truc, **Anh Nguyen To**, Tuan Ha Manh, Hien Ho Van, Tuong Nguyen Manh, Kien Trinh Trung, Khanh Truong Huu, Nhan Le Nguyen Thanh, Hung Nguyen Thanh, Chau Nguyen Van Vinh, Thwaites Guy, van Doorn H Rogier, and Tan Le Van, 2016. The first genome sequences of human bocaviruses from Vietnam. Wellcome Open Res. 1, 16.
10. Tu Minh, Hoang Van, **Anh Nguyen To**, Thanh Tan, Ty Thi, Vu Hang, Chau Vinh, Van Nguyen, Thwaites Guy E, Wills Bridget, Thwaites C Louise, Le Van Tan, and Doorn H Rogier Van, 2019. Clinical and aetiological study of hand , foot and mouth disease in southern Vietnam , 2013 – 2015 : Inpatients and outpatients. Int. J. Infect. Dis. 80, 1–9.
11. Van Hoang Minh Tu, **Anh Nguyen To**, Hong Nguyen Thi Thu, Nhu Le Nguyen Truc, Nguyet Lam Anh, Thanh Tran Tan, Ny Nguyen Thi Han, Hang VuThi Ty, Khanh Truong Huu, Viet Ho Lu, Viet Do Chau, Tuan Ha Manh, Hung Nnguyen Thanh, Quy Du Tuan, Ha Do Quang, Qui Phan Tu, Nhan Le Nguyen Thanh, Thwaites Guy, Chau Nguyen Van Vinh, Thwaites Louise, Van Doorn H Rogie, and Tan Le Van, 2019. Enterovirus A71 phenotypes causing hand, foot and mouth disease, Vietnam. Emerg. Infect. Dis. 25.

Abbreviations

&	And
µg	Microgram
µM	Micromolar
µmol	Micromole
µL	Microliter
kg	Kilogram
ADV	Human mastadenovirus
AFP	Alpha-fetoprotein
AIE	Autoimmune encephalitis
ALT	Alanine transaminase
BLAST	Basic Local Alignment Search tool
bp	Base pairs
CA	Community-acquired
Cap	Capsid protein coding sequences
cDNA	Complementary DNA
CDS	Coding sequences
CMV	Cytomegalovirus
CNS	Central nervous system
COVID-19	Coronavirus disease 2019
CSF	Cerebrospinal fluid
cu mm	Cubi millimeter
DALYs	Disability-adjusted life-years rate
ddNTP	Dideoxyribonucleotide triphosphate
DENV	Dengue virus
dL	Deciliter
DNA	Deoxyribose nucleic acid
dNTP	Deoxyribonucleotide triphosphate
ds	Double stranded
E30	Echovirus 30
EBV	Epstein-Barr virus
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
EVs	Enteroviruses
EV-A71	Enterovirus A71
EV-D68	Enterovirus D68
EEG	Electroencephalogram
FA	Influenza A virus
FB	Influenza B virus
GBD	Global Burden of Disease

GCS	Glasgow Coma Score
HAV	Hepatitis A virus
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HHV-6	Human hespervirus 6
HICs	High income countries
HIV	Human immunodeficiency virus
HPgV-2	Human pegivirus 2
Hr	Hour
HRV	Human rhinovirus
HSV	Herpes simplex encephalitis
HSV-1	Herpes simplex encephalitis type 1
HSV-2	Herpes simplex encephalitis type 2
HTD	Hospital for Tropical Diseases
ICU	Intensive care unit
JEV	Japanese encephalitis virus
IgG	Immunoglobulin G
IgM	Immunoglobulin M
Kb	Kilobase pair
Kg	Kilogram
kPa	Kilopascal
L	Liter
LMICs	Lower and middle-income countries
NA	Nucleic acid
ng	Nanogram
MAP	Mean arterial pressure
MCV	Molluscum contagiosum virus
MERS-CoV	Middle East respiratory syndrome coronavirus
mg	Miligram
min	Minute
ml	Mililiter
mmHg	Millimeters of mercury
mmol	Milimole
mM	Milimolar
mm^3	Cubic milimeter
mNGS	Metagenomic next-generation sequencing
MPV	Metapneumovirus
NCBI	National center for biotechnology information
NGS	Next generation sequencing
NMDAR	N-methyl-D-aspartate receptor
NS2/3	Non-structural 2-3 coding region
NS3	Non-structural 3 coding region

NS5B	Non-structural 5B coding region
PEV	Parechovirus
PIV 1	Parainfluenza virus 1
PIV 2	Parainfluenza virus 2
PIV 3	Parainfluenza virus 3
PIV 4	Parainfluenza virus 4
PCR	Polymerase chain reaction
qSOFA	Quick Sequential Organ Failure Assessment
Rep	Replication protein coding sequences
RNA	Ribonucleic acid
rPCR	Random PCR
RSV	Respiratory syncytial virus
RT-PCR	Real time PCR
SARS-CoV-1	Severe acute respiratory coronavirus 1
SARS-CoV-2	Severe acute respiratory coronavirus 2
SBP	Systolic blood pressure
SBS	Sequencing by synthesis
SD	Standard deviation
s	Seconds
SDI	Socio-demographic Index
SISPA	Sequence-independent single primer amplification
SIRS	Systemic Inflammatory Response Syndrome
SNP	Single nucleotide polymorphism
SOPA	Sequential Organ Failure Assessment
SRA	Sequence read archive
ss	Single strand
SURPI	Sequence-based ultra rapid pathogen identification
TTV	Torque teno virus
U	Enzyme unit
UK	United Kingdom
US	United State
US$	US dollar
UTR	Untranslated region
VIZIONS	Vietnam Initiative on Zoonotic Infections
VP1	Viral capsid protein 1
VZV	Varicella zoster virus
WBCs	White blood cells
WNV	West Nile virus
WHO	World Health Organization
YLD	Years lived with disability
ZIKV	Zika virus
ZMW	Zero Mode Waveguide

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1. Sepsis

Sepsis is the presence of systemic host responses to an infection. It is one of the leading causes of intensive care units admission of all age groups worldwide, and is considered as a serious public health problem (1). Sepsis is known as the final common pathway to death from severe infectious diseases (2).

1.1. Definition:

The definitions of sepsis and its diagnostic criteria have evolved over time. Major milestones have been the establishment of the first case definition in 1991, and its revised version in 2001, 2012 and 2014. According to the 1991 consensus conference, sepsis was defined based on the concept of the Systemic Inflammatory Response Syndrome (SIRS). Accordingly, sepsis was defined by the presence of two or more of SIRS including (1) temperature $>38^0$C or $<36^0$C; (2) an elevated heart rate (>90 beats/minute); (3) tachypnea (respiratory rate >20 breaths/minute or $PaCO_2$ <32 mmHg); and (4) an alteration in white blood cell count (>12,000/cu mm, <4,000/cu mm, or >10% immature forms) (3). Ten years later, the 2001 consensus conference proposed to expand the list of sepsis diagnostic criteria (4). The expanded diagnostic criteria were then adopted and recommended by the surviving sepsis campaign in subsequent years, especially in 2012 (Table 1.1). Most recently, the international conference was held in 2014 to deliver a new consensus definition for sepsis. According to the 2014 *Sepsis-3 Task Force*, sepsis is defined life-threatening organ dysfunction caused by a dysregulated host response to infection (5).

Table 1.1 Diagnostic criteria for sepsis. Table was adapted from "Surviving Sepsis Campaign: International Guidelines for Management of Severe Sepsis and Septic Shock: 2012" by Dellinger et al, 2013 (5).

General variables
Fever (> 38.3^{0}C)
Hypothermia (Core body temperature < 36^{0}C)
Heart rate > 90 beats per minute or more than 2 SD above the normal value for age
Tachypnea
Altered mental status
Significant edema or positive fluid balance (> 20ml/kg over 24hr)
Hyperglycemia (plasma glucose >140mg/dL or 7.7mmol/L) in absence of diabetes
Inflammatory variables
Leukocytosis (white blood cell count > 12,000/µL)
Leukopenia (white blood cell count < 400/µL)
Normal white blood cell count with greater than 10% immature forms
Plasma C-reactive protein more than 2 SD above the normal value
Plasma procalcitonin >2 SD above the normal value
Hemodynamic variables
Arterial hypotension (systolic blood pressure (SBP) < 90 mmHg, mean arterial pressure < 70 mmHg, or SBP decrease > 40 mmHg in adults or less than 2SD below normal for age)
Organ dysfunction variables
Arterial hypoxemia (PaO2 / FIO2 < 300)
Acute oliguria (urine output < 0.5 ml/kg/hr for at least 2hrs despite adequate fluid resuscitation)
Creatinine increase > 0.5 mg/dL or 44.2 µmol/L
Coagulation abnormalities (international normalised ratio >1.5 or activated partial thromplastin time >60 seconds)
Ileus (absent bowel sounds)
Thrombocytopenia (Platelet count < cells 100,000/µL)
Hyperbilirubinaemia (plasma total bilirubin > 4 mg/dL or 70 µmol/L)
Tissue perfusion variables
Hyperlactatemia (> 1 mmol/L)
Decreased capillary refill or mottling

The organ dysfunction is assessed by the Sequential Organ Failure Assessment (SOFA) scores (Table 1.2). A patient with a SOFA score of 2 or more is considered to have sepsis and is associated with a mortality rate of approximately 10%. The higher SOFA score is associated with a greater mortality risk. qSOFA (quick SOFA) was introduced by *Sepsis-3* as a simple, quick assessment without the need of laboratory test at bedside for adult patients with suspected infection, who are likely have poor outcome (6).

Table 1.2 Sequential Organ Failure Assessment Score. The table was adapted from "The third international consensus definitions for sepsis and septic shock" by Singer et al, 2016 (6).

System	Score				
	0	1	2	3	4
Respiration					
PaO2/FIO2, mmHg (kPa)	≥400 (53.3)	<400 (53.3)	<300 (40)	<200 (26.7) with respiratory support	<100 (13.3) with respiratory support
Coagulation					
Platelet, x 10^3 /µl	≥ 150	<150	<100	<50	<20
Liver					
Bilirubin, mg/dL (µmol/L)	<1.2 (20)	1.2-1.9 (20-32)	2.0-5.9 (33-101)	6.0-11.9 (102-204)	>12.0 (204)
Cardiovascular	MAP ≥ 70 mmHg	MAP < 70 mmHg	Dopamine <5 or dobutamine (any dose)	Dopamine 5.1-15 or epinephrine ≤0.1 or norepinephrine ≤0.1	Dopamine >15 or epinephrine >0.1 or norepinephrine >0.1
Central nervous system					
Glasfow Coma Scale score	15	13-14	10-12	6-9	<6
Renal					
Creatinine, mg/dL (µmol/L)	<1.2 (110)	1.2-1.9 (110-170)	2.0-3.4 (171-299)	3.5-4.9 (300-440)	>5.0 (440)
Urine output, ml/d				<500	<200

According to Sepsis-3 task force, septic shock is defined as a subset of sepsis in which underlying circulatory and cellular/metabolic are profound enough to substantially increase mortality (5). More specifically, septic shock is defined by the clinical criteria of sepsis with the persistence of hypotension requiring vasopressor to maintain the mean arterial pressure of 65mmHg or above, and a serum lactate greater than 2 mmol/L (18 mg/dL) despite adequate fluid resuscitation (6). The hospital mortality among patients with a sepsis shock is >40% as compared to ≈10% among those with sepsis.

It should however be noted that the definition of sepsis has largely been based on patients with bacterial or fungal infection. So far, the definition has not validated for other infections such as malaria, dengue or other viruses (7).

1.2. Global burden:

The true global epidemiological burden of sepsis has not been fully documented. Available epidemiological data have been derived from hospital-treated sepsis data in high-income countries (8). There is a severe lack of population-based sepsis data globally. The incidence of sepsis is less well described in the low and middle-income countries (LMCs), where 87% of the world's population lives (2). A recent Global Burden of Disease (GBD) study conducted across 195 countries and territories estimates global incidence of sepsis was 48.9 million (38.9–62.9) cases, and a total of 11 million sepsis-related deaths were reported worldwide in 2017, representing 19.7% of all deaths that year (Figure 1.1)(9).

The incidence of sepsis peaks in early childhood and again in elderly adults. About 41.5% (20.3 million) of sepsis cases and 26.4% (2.9 million) deaths related to sepsis worldwide are among children younger than five years (9). According to global age-standardised estimation, in 2017 sepsis incidence was higher among females than males (716.5 cases per 100,000 vs. 642.8 cases per 100,000), while sepsis-related mortality was higher among males than females (164.2 per 100,000 vs. 134.1 per 100,000). In the same year, diarrheal disease is the largest contributor of sepsis incidence (9.2 million), whereas the highest number of sepsis-related deaths were caused by lower respiratory infections (1.8 million) (9). Of these, 70% was community-acquired (CA) sepsis (8). The intensive care unit (ICU) and hospital mortality rates of patients with sepsis were significantly higher than those in the general population (25.8% vs. 16.2%; 35.3% vs. 24.2%, respectively) (10).

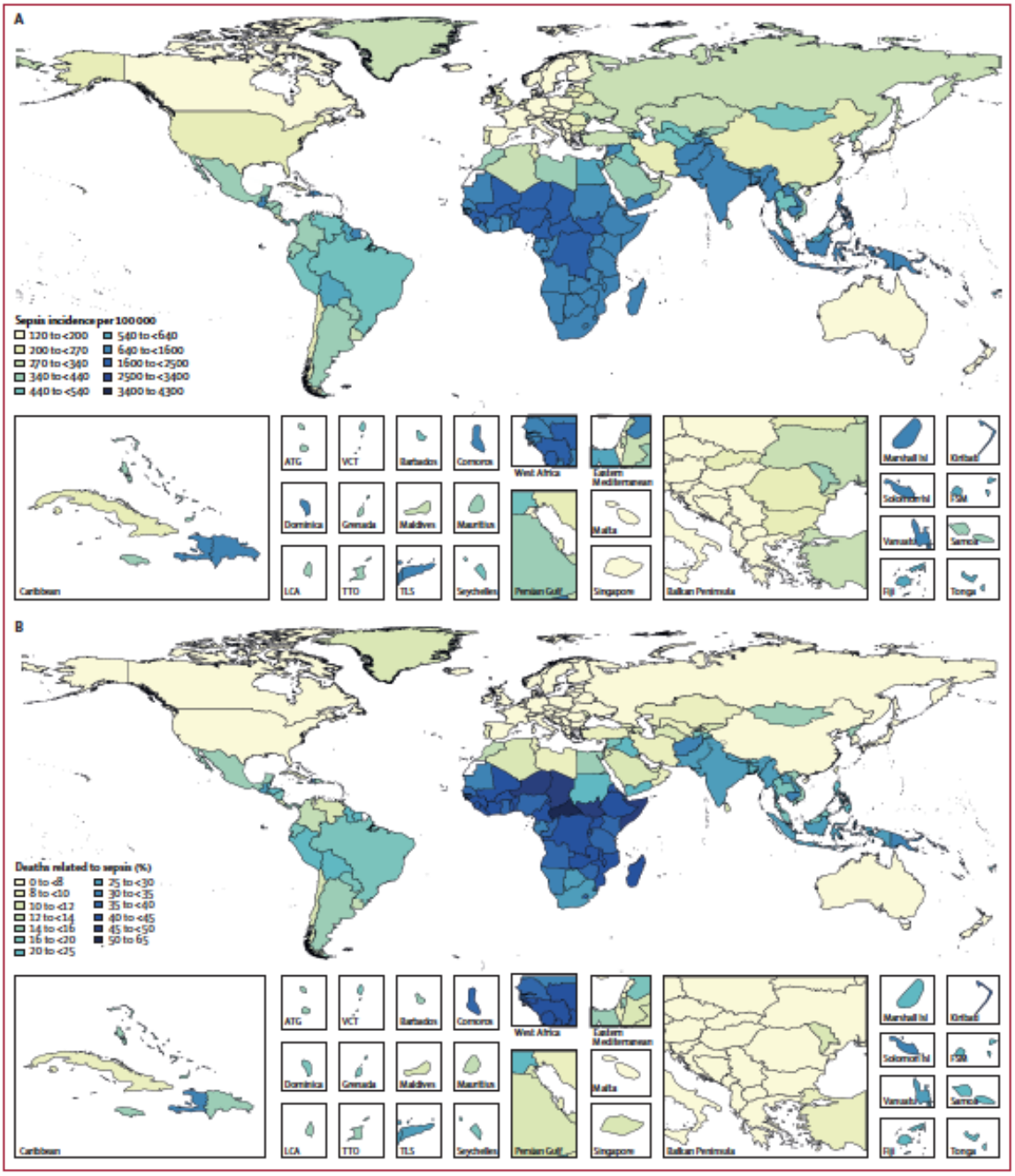


Figure 1.1 Age-standardised sepsis incidence per 100,000 population for both sexes, in 2017 (A), and percentage of all deaths related to sepsis, age-standardised for both sexes, in 2017 (B) ATG=Antigua and Barbuda. FSM=Federated States of Micronesia. LCA=Saint Lucia. TLS=Timor-Leste. TTO=Trinidad and Tobago. VCT=Saint Vincent and the Grenadines. The figure was adapted from "Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study" by Rudd et al, 2020 (9).

In high income countries (HICs), sepsis accounts for approximately 2% of all hospitalizations (11). A meta-analysis study of 27 studies from 7 high-income countries estimated that the incidence rates of hospital-treated sepsis and severe sepsis were 437 and 270 cases per 100,000 people during 2003-2015, respectively. The mortality rates were 17% and 26% for sepsis and severe sepsis, respectively (12). In the US, the National Center for Health Statistics reported that over a period of 8 years (2000 – 2008) the annual incidence of sepsis rose by 7 to 8% per year (13), and more than 200,000 persons died of sepsis in 2007 (14). In Germany, the average annual incidence of sepsis cases increased by 5.7% over a period of 6 years (2007-2013). Meanwhile, the hospital mortality of sepsis and severe sepsis fell from 27% to 24.3% and from 49.5% to 43.6%, respectively (15). The reported annual incidence of severe sepsis increased by 8.6% in Spain (16) and 43% in the United Kingdom during 1996-2004 (17). Elsewhere in Australia and New Zealand, of 12,512 critically sick patients, the reported mortality was 20% (18).

Scarce data exists regarding the epidemiology of sepsis in LMICs. In these settings, the burden is apparently higher than that in the developed countries. Notably, the GDB study estimated that around 85% of sepsis cases and 84.8% of related deaths worldwide occurred in LMICs, particularly in sub-Saharan Africa and South-East Asia (9). In Malawi, a recent report showed that the estimated incidence rates of emergency department attending sepsis and severe sepsis in adults between 2013 and 2016 were estimated to be 1772 and 303 cases per 100,000 person-years, respectively (19). The highest incidence was observed in the group of oldest patients (≥ 80 years of age). Meanwhile, the estimates of fatality rates of sepsis and severe sepsis were 23.7% and 28.1%, respectively, and the rate increased with age (19). A prospective observational study conducted in a tertiary care hospital in Northeast Thailand during 2013-2017 reported that about 13% (3,716/28,752) of the screened patients had evidence of sepsis

and 21% of these patients died (20). In Indonesia, a very high mortality rate of sepsis reported was 58.3% (21).

More than 50% of patients with severe sepsis would require intensive care services (22). Therefore, sepsis causes significant economic burden as it is costly and tends to consume a lot of resources. A systematic review of hospital-related costs of sepsis based mostly on data from HICs showed that the median of the mean hospital-wide cost and ICU costs of sepsis were approximately more than US$ 32,000 and US$ 27,000 per patient per stay, respectively (23). In the US, sepsis treatment is listed as the most expensive condition in hospitals (US$ 20 billion annually) (24). However, measuring the precise costs of sepsis treatment was complicated due to the heterogeneity in the definition and nature of the disease.

1.3. Etiology:

A broad range of pathogens (bacteria, fungi and viruses) can cause sepsis, while many infections (respiratory, central nervous system or enteric infections) may lead to sepsis. The causative organisms of sepsis have evolved over time. This is in part attributable to the emergence of novel pathogens, such as SARS-CoV-2, and the implementation of vaccination programmes globally. Available data have shown that respiratory infection is the most common source of sepsis followed by genitourinary and abdominal sources (25,26). However, nearly 50% of sepsis patients had no etiological agent identified (26–28).

1.3.1. Bacteria:

Bacteria are the most common causes of sepsis. Historically, sepsis was regarded as a disease related to gram-negative bacteria infection (29). However, recent studies demonstrated that gram-positive bacteria were as common as gram-negative bacteria in causing sepsis (30,31). The most common gram-positive bacteria isolated from sepsis patients were *Staphylococcus aureus* and *Streptococcus pneumoniae*, whereas *Escherichia coli*, *Klebsiella* species, and

Pseudomonas aeruginosa were predominant gram-negative bacterial pathogens of sepsis (26,32,33). Moreover, the increasing antimicrobial resistance has greatly impacted the management of sepsis. As a consequence, drug-resistant pathogens cause higher mortality (34,35). Important emerging multi-drug resistant pathogens include methicillin-resistant *S. aureus*, vancomycin-resistant enterococci, extended-spectrum beta-lactamase-producing gram negatives, and multiresistant strains of *P. aeruginosa* and Acinetobacter species.

1.3.2. Fungi:

Sepsis associated with fungal infection has been reported with increasing frequency, causing substantial morbidity and mortality (36). Fungal infections are responsible for around 5% of sepsis cases (30,31,37). *Candida* species is the most frequent causative agent of severe sepsis or septic shock in ICU patients (30,36,37).

1.3.3. Viruses:

Historically, sepsis caused by viral infection was neglected in part due to the lack of diagnostic capacity. Yet, a wide range of viruses, including emerging viruses such as SARS-CoV-2, can cause sepsis. Previous studies showed around 1% of sepsis cases were associated with viral infection (38,39). Most recently, an etiological study of sepsis conducted in Southeast Asia using a comprehensive panel PCR targeting at multiple viruses demonstrated that viruses were responsible for 29% of 1582 sepsis patients (26). Of the detected viruses, dengue virus was most predominant virus, accounting for 7.7% of sepsis cases. Dengue viruses (DENV) belong to the family *Flaviviridae* (40), and is a well-known cause of viral infection in tropical countries. The global estimated incidence of DENV is around 390 million infections per year (41). Reported data from Thailand showed that about 14% of patients with community acquired (CA) sepsis were tested positive for dengue viruses by PCR (42).

Herpes simplex virus and enteroviruses are the most common viral causes of neonatal sepsis (43). Among young children, enteroviruses and human parechoviruses are the most common causes of viral sepsis (44).

Respiratory infections are the major source of sepsis. A recent retrospective study included hospitalised patients showed that sepsis was documented in 61% of the patients presenting with viral CA pneumonia (45). Respiratory viruses can cause sepsis include influenza virus, rhinovirus, respiratory syncytial virus, parainfluenza virus types 1-3, and adenovirus. A retrospective cohort study from the US revealed that the incidence rate of influenza-associated severe-sepsis was 8.8 per 100,000 person-years (46). Additionally, severe sepsis was documented in 73% of hospitalizations attributable to influenza-associated critical illness (46). Infections with novel respiratory viruses such as SARS-CoV-1, avian influenza A virus subtype H5N1, MERS-CoV, and SARS-CoV-2 can lead to sepsis with high mortality (47–49). A report from China early this year showed that of 191 patients with confirmed SARS-CoV-2 infection, 59% and 20% had sepsis and septic shock, respectively with the median time from illness onset to sepsis of 9 days (range: 7.0–13.0 days) (50).

2. Central nervous system infections:

2.1. Definition:

Infectious diseases involving the central nervous system (CNS) have long been recognized as the most serious diseases with devastating clinical consequences that affect millions of people around the world. According to the WHO, they constitute the sixth cause of neurological consultation in primary care services (51). Significant morbidity and mortality often occur if patients with CNS infection are not recognized and promptly treated (52). The classification of CNS infections is based on the site of infections, with meningitis and encephalitis being major clinical entities. Meningitis and encephalitis will be further discussed herein.

2.1.1. Meningitis:

The term "meningitis" is used to define the inflammation of the leptomeninges and subarachnoid space, and is a neurologic emergency (53). Bacterial meningitis is an inflammation of the meninges affecting the pia, arachnoid, and subarachnoid space that happens in response to bacteria and bacterial products (54). A case of bacterial meningitis is defined with sudden onset and fever, intense headache, nausea, vomiting, and neck stiffness. In some specific circumstances, such as meningococcal disease, petechial rash with pink macules can be observed. A confirmed meningitis case requires a combination of clinical diagnosis and laboratory evidence demonstrating the presence of the respective pathogen in the cerebrospinal fluid (CSF) sample (51).

Bacterial meningitis is associated with serious morbidity (55). The outcomes following meningitis include seizure disorder, and motor and cognitive deficits (51). Bacterial meningitis is listed as a top 10 leading cause of death among communicable diseases (56).

In contrast to bacterial meningitis, viral meningitis is usually much less severe. It accounts for the majority of viral CNS infections cases (53). The inflammation occurs when viruses reach the meninges from the bloodstream or reactivate from a dormant state within the nervous system (57). Viral meningitis is usually self-limited without sequelae, especially in immunocompetent patients (58).

A case of viral meningitis is defined with an acute onset of meningeal symptoms, fever and CSF pleocytosis with no growth on routine bacterial culture (59). In adults, the general clinical features of viral meningitis include an acute onset of headache, neck stiffness, photophobia, and often nausea and vomiting, which are similar with those of bacterial meningitis (60,61). However, clinical features in children are often nonspecific depending on the child's age and the duration of illness (57). Seizure disorder is less frequently observed in pediatric patients with

viral meningitis (62,63). Clinically, bacterial and viral meningitis are often difficult to distinguish.

2.1.2. Acute encephalitis:

Acute encephalitis is considered a serious neurological condition, of which viruses are the most important and common pathogens. Encephalitis is an inflammation of the brain parenchyma due to direct brain invasion of neurotropic viruses through the blood–brain barrier or by retrograde axonal transport means (64). Sometimes, the inflammation can be caused by an indirect immunologically mediated injury accounting for approximately one-third of acute encephalitis cases (65).

Unlike viral meningitis, viral encephalitis can be life threatening and results in permanent neurological disability in both adults and children (58). Patients with viral encephalitis may present with a wide range of clinical features include an altered level of consciousness, fever, headache, psychiatric symptoms, cognitive defects, seizures, and focal neurologic deficits (66,67). A consensus definition of a viral encephalitis case was suggested by the International Encephalitis Consortium meeting in Atlanta, US in 2012. The major criterion is the alteration of mental status (defined as decreased or altered level of consciousness, lethargy or personality change) lasting for 24 hours or more with no alternative cause identified. Minor criteria (two for possible and at least three for probable or confirmed) include documented fever ≥38°C within 72 hours, seizures, new focal neurologic findings, CSF pleocytosis (≥5 white blood cells [WBCs]/μL), neuroimaging with brain parenchymal abnormality or electroencephalogram (EEG) consistent with encephalitis. A confirmed encephalitis case requires the demonstration of the presence of an infectious pathogen in brain biopsy or the detection of the pathogen genetic material or specific antibodies in the CSF (68).

Autoimmune encephalitis (AIE) is a newly recognized non-infectious encephalitis form. The principle of AIE is an immune response against neuronal auto antigens with production of antibodies against neuronal cell-surface or synaptic proteins (69). It is estimated the AIE accounts for some 21% of all encephalitis cases in England (70). First recognized in 2007 (71), anti-N-methyl-D-aspartate receptor (NMDAR) encephalitis is considered as the most common causes of AIE. AIE has a wide variety of clinical manifestations resembling other infectious encephalitis (72). AIE, especially anti-NMDAR encephalitis, occurs more frequently in children and young female patients (73).

2.2. Burden, epidemiology and incidence of CNS infections:

CNS infections affect millions of people around the world. The diseases are associated with high morbidity and mortality, thus posing substantial economic burden for individuals, families and the health care systems worldwide (51). According to the Global Burden of Disease (GBD) study 2015, meningitis and encephalitis globally caused 379,000 and 150,000 deaths respectively in 2015 (74). As a consequence, together with neurologic injuries, and nutritional deficiencies and neuropathies category, CNS infections are listed as one of the three categories that have the highest YLD (years lived with disability) in low income countries (51). According to global age-standardised estimation, CNS infection disorders were the largest cause of DALYs (disability adjusted life years) at low levels of SDI (Socio-demographic Index) (74).

According to the GBD study of meningitis 2016, the global incidence rates of meningitis increased from 2.50 million in 1990 to 2.82 million in 2016, while the global death rate decreased by 21% from 1990 (403,012 deaths) to 2016 (318,400 deaths) (75). Although the global mortality has declined, substantial disparities in geographical distribution and age groups persist. The incidence peaked during the neonatal period, and death rate is highest in children younger than 5 years (75).

The estimated number of people affected by bacterial meningitis worldwide was 2,907,146 each year. The highest incidence was observed in Africa, with 65 cases per 100,000 people, and the lowest incidence was recorded in the US or Canada, with 2/100,000 (76). The overall average of fatality was 15.9% (76), with the highest rate in Swaziland (32.7%) (77) and the lowest rate in Singapore (2.4%) (78). In developed countries, bacterial meningitis incidence rate is approximately 3 cases per 100,000 persons, while this rate is at least 10-times higher in developing countries (79). The variation of incidence, morbidity and mortality rates depend on several factors including age, geographical region, causative organism and immunization status (80). In some countries, bacterial meningitis is common in children younger than four years, with a peak in those aged 3–8 months (81). Vaccination targeting the major bacterial pathogens (such as *Haemophilus influenzae*, *Neisseria meningitidis*, *S. pneumonia*) have successfully reduced the incidence of meningitis in Africa, the Americas, Asia, Australasia, and Europe (75). However, due to the low or a drop in vaccination coverage in some low-income countries, the disease still affects a large number of children, especially those in the resource-constrained settings (82).

The estimated incidence of viral CNS infections is 20–30 per 100,000 per year (62), roughly three times as common as bacterial infections (83). The incidence of viral meningitis is estimated to range from 0.26 and 17 cases per 100,000 people (57). The incidence rate tends to decrease with age, presumably attributable to the immune status. Many studies reported the highest incidence rate in young children (84,85). In countries where there is a high immunization coverage for bacterial pathogens, viruses are major causes of meningitis (57). In the US, an average hospitalization incidence of viral meningitis was around 36,000 persons per year (14 per 100,000 population), which resulted in an annual estimated cost of US$ 234-310 million during 1988-1999 (86). Viral meningitis commonly presents in the summer and autumn months

in temperate climates, while it exhibits a year-round incidence in areas with tropical and subtropical climates (84,87,88).

Despite the high morbidity and mortality, the magnitude of epidemiology of viral encephalitis is understudied due to the lack of harmonized definition and the variations in diagnostic criteria (89). The global incidence of viral encephalitis is estimated around 3.5–7.4 cases per 100,000 patient-years. Of these, the incidence of pediatric encephalitis is more than 16 cases per 100,000 patient-year (65,90). The fatality of viral encephalitis ranges from 4.6% to 29% (91). In the western world, the reported incidence of encephalitis ranges from 0.7 to 13.8 per 100,000 for all ages; approximately 0.7 – 12.6 per 100,000 in adults, and 0.7 – 12.6 per 100,000 in children (92). In the US, the average annual hospitalization incidence rate of viral encephalitis was 6.9 per 100,000 persons with the fatal rate of 5.8% during 1998-2010. The median cost of an encephalitis-associated hospitalization was US$ 23,518 in 1998 and US$ 48,852 in 2010 (93).

In Vietnam, the median incidence of bacteria and non-bacteria CNS infections in children during 2010-2015 were 64.5 and 51.4 per 100,000 district population in southern Vietnam, respectively (94). High morbidity and mortality were reported in both the north and south of Vietnam; mortality rate was 8-12% and residual disability rate was 10-30% (95,96).

2.3. Viral etiology of CNS infection:

There are many infectious organisms that have been recognized to cause CNS infections including broad categories of bacteria, viruses, fungi, mycobacteria and parasites. Yet, around >60% of patients presenting with CNS infections had no etiology identified (70,93,99–101). Likewise, previous studies from Vietnam failed to identify a causative agent in more than half of the patients (91,95,100,101). The non-viral pathogens are listed in Table 1.3 and 1.4. More than 100 viral pathogens have been recognized to cause CNS infections (65). Because of the

focus of the present PhD research, here I will focus on common viral causes of CNS infections (Table 1.5), especially those circulating in Asia.

Table 1.3 Common non-viral pathogens cause meningitis

Bacterial/mycobacterial pathogens	*Streptococcus pneumonia, Haemophilus influenzae*, *Neisseria meningitidis, Listeria monocytogenes, Group B streptococcus, Escherichia coli, Mycobacterium tuberculosis*.
Fungal pathogens	*Cryptococcus neoformans, Coccidioides immitis, Blastomyces dermatitidis, Histoplasma capsulatum, Candida spp, Aspergillus spp, Zygomycetes*
Parasites	*Strongyloides stercoralis, Naegleria fowleri, Angiostrongylus cantonensis*

Table 1.4 Other organisms can cause meningitis. Table was adapted from "Epidemiology of Central Nervous System Infection" by Riddell et al, 2012 (52).

Pathogen	Risk Factor	Incidence
Borrelia burgdorferi (Lyme)	Regional tick exposure	Regional
Treponema pallidum (syphilis)	Sexual transmission	Common with HIV coinfection
Leptospirosis	Contaminated fresh water	Regional
Rickettsia rickettsii	Tick borne	Regional
Rickettsia typhi	Fleas	Regional
Rickettsia prowazekii	Lice	Regional, epidemic
Rickettsia tsutsugamushi	Mites	Regional
Ehrlichia spp	Tick borne	Regional
Bartonella henselae	Cat-scratch	Rare
Coxiella burnetii	Contact with infectious material from infected animals	Rare

Table 1.5 Common viral causes of Central Nervous System infections. Table was adapted from "Management of Viral Central Nervous System Infections: A Primer for Clinicians" by Bookstaver et al, 2017 (102).

Viral type	Pathogen
dsDNA	*Herpes simplex virus, Varicella zoster virus, Epstein-Barr virus, Cytomegalovirus, Human hespervirus 6*
(+) ssRNA	*Enteroviruses, Dengue virus, Japanese encephalitis virus, West Nile virus, Zika virus, Rubella virus, tick-borne encephalitis virus, Murray Valley encephalitis virus, St Louis encephalitis virus.*
(-) ssRNA	*Rabies virus, Measles virus, Mumps virus, Nipah virus, Hendra virus*

2.3.1. Enteroviruses:

Enteroviruses (EVs) are regarded as the most common pathogens of viral meningitis. EVs are non-enveloped single-stranded RNA viruses of the family *Picornaviridae*. EVs account for 23% to 61% of meningitis cases with a pathogen identified, especially in young children (59).

There are more than 110 genetically distinct human EV serotypes that have been identified and known to cause human infection. Echovirus 6 and 30 are responsible for the majority of meningitis cases caused by EVs (103–106). Less common causes include enterovirus A71, echovirus 9, 13, 14 and 16, coxsackievirus A9 and B5 (59). Enterovirus A71 infection can lead to brain stem encephalitis, which can be fatal. The circulation of enterovirus A71 in Asia has triggered 2-3 year cycles of hand foot and mouth disease outbreaks associated with severe condition such as encephalitis and high mortality (107,108). In recent years, enterovirus D68 has emerged and has been linked with CNS infections, especially acute flaccid myelitis in young children (109).

Infants and young children with no immunity are most susceptible to EVs, and the incidence decreases with age (110). The peak of EV infections falls in late summer and autumn (111). Infection occurs by the fecal-oral route, followed by viral replication in the gastrointestinal tract and subsequent viral dissemination into the blood and occasionally the CNS. Most EVs meningitis cases are self-limited and have a good prognosis. However, considerable morbidities could occur with moderate or high fever despite the use of antipyretics and several days of severe headache warranting opiate analgesia (112).

2.3.2. Herpes simplex virus:

Herpes simplex encephalitis (HSV) is a member of the family *Herpesviridae*. HSV is the most important cause of viral encephalitis worldwide. If untreated, the mortality of HSV associated encephalitis can be up 70%, and a substantial proportion of the survivors still suffer from long

term severe sequelae (51). Sporadic HSV encephalitis is caused by HSV type 1 (HSV-1), commonly occurs in patients younger than 20 or those older than 50 years of age (111). In the US, the annual incidence of HSV encephalitis is 1 case per 250,000 population (113). In Sweden, the reported incidence is 2.2 cases per million population of HSV-1 related encephalitis per year (113).

HSV type 2 (HSV-2) is associated with benign recurrent aseptic meningitis and recurrent benign lymphocytic meningitis. HSV meningitis accounts for 0.5% to 18% of viral meningitis cases (59). HSV-2 is a common sexually transmitted infection associated with oral and genital mucocutaneous lesions (114).

Other latent viruses belong to the family *Herpesviridae* that can cause CNS infections due to reactivation are Varicella zoster virus (VZV), Epstein-Barr virus (EBV), Cytomegalovirus (CMV) and Human hespervirus 6 (HHV-6). A previous report showed VZV was responsible for 8% of meningitis and 12% of encephalitis in adults (84). Although EBV infection were detected in 25% of meningitis and 43% of encephalitis cases in a retrospective descriptive study in Qatar from 2011-2015 (115), EBV and CMV may not be directly linked with the ongoing CNS infections.

2.3.3. Rabies virus:

Rabies is one of the oldest and the most feared diseases documented in the medical literature. Rabies viruses are single negative-stranded RNA viruses belonging to the genus *Lyssavirus* of the family *Rhabdoviridae*. Rabies is a viral zoonotic disease, with dogs being the most common reservoir. Infection with the rabies virus can result in severe encephalitis with the mortality of almost 100%. There has been only one report of a single case surviving from rabies to date who did not receive any rabies vaccines prior her presentation (116). It is estimated that 59,000 persons die of rabies each year, of which more than 95% occurs in Africa and Asia (117).

However, rabies virus infection is preventable by vaccination. In the US where canine vaccination coverage is high, the incidence of rabies has declined substantially from 33 cases in 1943 to 3 cases in 2006 (118).

2.3.4. Flaviviruses:

Flaviviruses are single-stranded RNA viruses, and belong to the family *Flaviviridae.* Encephalitis-associated flaviviruses (dengue virus, Japanese encephalitis, West Nile virus, and Zika virus) are geographic dependent (Figure 1.2). Figure 1.2 shows global distribution of common flaviviruses causing CNS infections. Flaviviruses are responsible for a significant proportion of CNS infections worldwide.

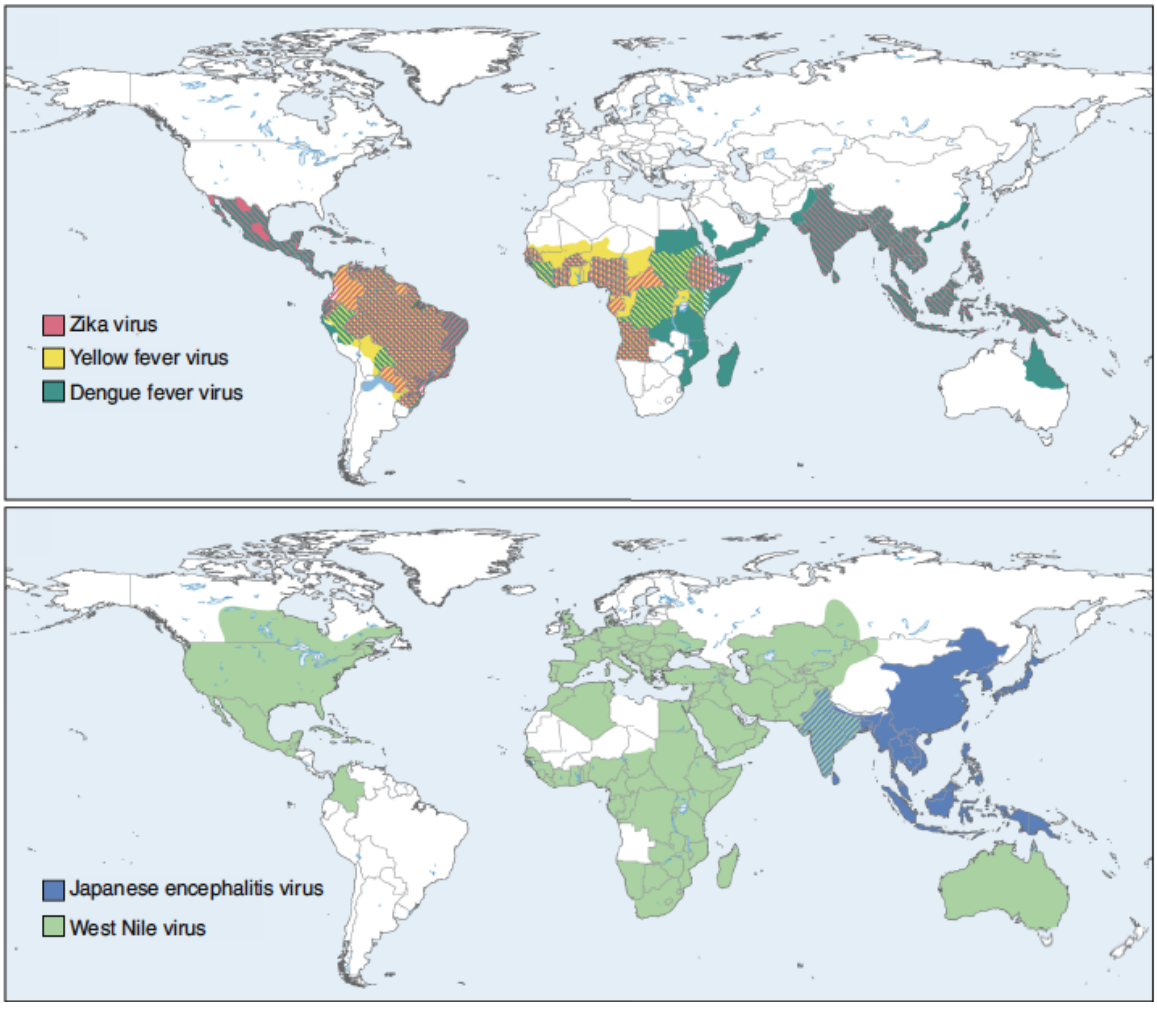


Figure 1.2 Global distribution of common flaviviruses causing CNS infection. Figure was adapted from "The continued threat of emerging flaviviruses" by Pierson et al, 2020 (119).

Dengue virus

There are four serotypes of dengue viruses (DENV) (1-4). DENV is transmitted to human via the bites of *Aedes mosquitoes* but mostly *Ae. Aegypti* mosquitoes carrying the virus (119). According to WHO, the number of reported dengue cases globally in 2019 was 4.9 million (120). Meningitis or encephalitis is a rare clinical condition of DENV infection, and is more frequently observed in South, Central America and Southeast Asia where the prevalence of DENV is high.

Japanese encephalitis virus:

Japanese encephalitis virus (JEV) is transmitted by *Culex* mosquitoes, particularly *Culex tritaeniorhynchus.* Pigs serve as an amplifying host of the zoonotic cycle, of which humans is the dead end host (119). The estimated annual incidence and death of JE are 67,900 and 13,600–20,400, respectively with the majority of the cases occurring in Asia (121). JEV is endemic in Asia with annual peaks occurring during late summer and fall (58). Despite the availability of effective vaccines, JEV remains an important causes of encephalitis in children in Asia, including Vietnam (65).

West Nile virus:

West Nile virus (WNV) can caused serious outbreaks of meningoencephalitis and has established its endemic cycles in the US since its first arrival in there in 1999. WNV is transmitted to humans by *Culex* mosquitoes from its natural host, birds. WNV is primarily transmitted by the Culex species mosquito (111). After the report of the first 59 confirmed cases in New York city in 1999, WNV caused a widespread epidemics across the US along bird-migratory routes during the summer of that year (122). In 2002, the reported number of neuroinvasive cases related to WNV was 2942, accounting for 71% of the cases of viral meningoencephalitis reported in the US (123). In 2009, there were only 373 cases of encephalitis

or meningitis reported in the US (124), and over a 10-year period, the fatality rate of WNV disease was about 10% (58).

Zika virus:

Zika virus (ZIKV) is an emerging virus and is transmitted by daytime-active *Aedes* mosquitoes. ZIKV was first discovered in 1947 in Africa. The virus was not detected in the Western hemisphere until 2015 (58). In 2015, the first case of ZIKV infection was recognized in Brazil. In subsequent months, the virus rapidly spread through the region and many other parts of the world. It was estimated that over 1.3 million ZIKV cases were recorded by the end of 2015 in Brazil alone (125). The geographic distribution of ZIKV has now been expanded to sub-Saharan Africa, most of Central America, South America, and the Caribbean, and the United States (125). Infection with ZIKV during pregnancy may lead to microcephaly in the newborn (126). The risk of ZIKV infection to the fetus has been understudied (127).

2.3.5. Measles viruses:

Measles virus is a single negative-stranded RNA virus of the genus *Morbillivirus* and the family *Paramyxoviridae*. Measles is highly contagious exanthematous viral infection with 20 million cases reported annually worldwide (128). Clinical features of measles includes a combination generalized rash lasting in 3 days or more, fever, and cough, coryza, or conjunctivitis (58). Measles infection occurs by inhalation of aerosolized droplets. Viremia can occur when local dendritic cells pick up the virions and carry to lymph nodes (129) . Acute meningoencephalitis caused by measles infection is rare, about 0.1% of the infected cases, with a mortality of between 10% and 15%, and one-fourth of the survivors suffered permanent neurologic sequelae (129).

Subacute sclerosing panencephalitis (SSPE) is a fatal neurological complication associated with persistent measles infection, which might occur from 4 to 10 years after an acute measles

infection, and the latency period varies from 1 month to 27 years (130). The intracerebral spread of measles virus leads to destruction of neurons (131). SSPE is common in children and young adults with high mortality but no specific treatment available (132). The incidence of SSPE in the US and Israel are from 4 to 11 per 100,000 and 23 per 100,000 measles cases, respectively (130). Measles can be eliminated by routine vaccination of children, however outbreak can still occur due to the importation of unvaccinated cases (133) or a drop in vaccination coverage. A recent outbreak of measles in Vietnam in 2014 caused some 15,033 confirmed cases and 140 deaths (134).

2.3.6. Mumps and Rubella viruses:

Mumps virus (MuV) was first isolated in 1945. It belongs to the genus *Orthorubulavirus* of the subfamily *Rubulavirinae* and the family *Paramyxoviridae*. Mumps virus is highly contagious and the infection can lead to meningitis. During the 1940s and 1950s, MuV meningitis accounted for 15% of cases presenting with viral meningitis (88).

Rubella virus (RuV) is a single positive-stranded RNA virus, which is a member of the genus *Rubivirus* and the family *Matonaviridae*. Neurological complications caused by RuV infection are known as the congenital rubella syndrome with significant parenchymal volume loss in the brain and a static encephalopathy (58). In 1996, an estimated 22,000 babies in Asia, 46,000 in South-East Asia and 13,000 in West Pacific were born with the congenital rubella syndrome (135).

Both MuV and RuV are preventable through vaccination. As such, the incidence of mumps and rubella has significantly been reduced over the last decades because of the vaccination programme worldwide. Meningitis associated with MuV infection is rare, especially in developed countries nowadays (136). Likewise, acute encephalitis related to RuV is also rare,

accounting for 0.02% of the reported encephalitis cases but the mortality can be up to 20% (135).

2.3.7. Nipah and Hendra viruses:

Hendra and Nipah virus are recently discovered zoonotic viruses that can cause encephalitis. Genetically, they are two closely related paramyxoviruses but have different geographic distribution. Nipah virus was first discovered during an outbreak of encephalitis of unknown origin occurring in Malaysia and Singapore in 1998 among abattoir workers (137). Subsequently, pigs were then recognized as the sources of the transmission, and bats were the natural reservoirs of Nipah virus (137). After 1998, Nipah virus has been reported to cause human infections in India and Bangladesh. Hendra virus was discovered in 1994 when it caused deaths of a horse trainer and many horses in Australia in 1994 (138). To this end, Hendra virus infection has only been reported in Australia (139).

3. Laboratory diagnosis of sepsis and CNS infections:

Rapid and accurate detection of the causative agents is crucial for timely clinical intervention, thereby improving outcome in patients with sepsis or CNS infections. Bacterial culture of blood and CSF samples are considered as the gold standard for sepsis diagnosis and CNS infections. The method detects viable microbial pathogens in the tested samples, therefore allows for subsequent antimicrobial susceptibility assessment. The results are critical to initiate appropriate antimicrobial therapy (140). However, bacterial culture has several limitations including long turnaround time and low sensitivity, especially in patients with prior use of antimicrobials. A routine culture procedure could take 6h to 5 days or even longer to complete, including the time required for species identification (24h) and antimicrobial susceptibility testing (48h) (141). About 25-50% of sepsis patients had a negative culture result (142). Contributing factors include blood volume, microbial titers in the tested samples and the prior use of antimicrobial agents

(143,144)(145). Meanwhile, microscopic examination of CSF or blood samples can tell whether there is a presence of a gram negative or positive bacterium in the samples under investigation. The sensitivity of gram stain for diagnosis of bacterial meningitis is also highly dependent on the bacterial loads and the prior use of antimicrobials of the patients. Under an ideal circumstance, the diagnostic yields of gram stain can reach 70%-90% in patients without antimicrobial treatment and 40%-60% in patients with antimicrobial treatment (146). Because of the focus of the current thesis study, I will focus my literature review more on the methods commonly used for virus diagnosis.

3.1. Virus culture:

Virus cultures from clinical specimens are usually performed in common cell lines such as African green monkey kidney (Vero) cells, human amniotic epithelial cells and human embryonic skin fibroblasts (147). The inoculation is subjected to daily inspection for cytopathic effect. Immunofluorescence test or nucleic acid based detection methods are often used for detection of the virus causing cytopathic effect in culture materials. A positive culture result of brain tissues offers a definitive diagnosis of CNS infections (148). However, collecting biopsies are highly invasive and are not allowed in most of the settings. CSF is thus the useful sample for routine diagnosis of CNS infections. While the specificity of virus isolation of CSF samples can be up to 100%, its sensitivity is usually poor because of late admission and/or the lack of appropriate culture cell lines that can recover all possible viruses causing CNS infections. The sensitivity of CSF culture of EVs is however approximately 65-75% (149). The most common cell lines used for EVs culture are Vero cells. In addition to the low sensitivity, virus culture is time-consuming. It may take up to 6 days or more to complete a virus culture experiments (150). This might delay the initiation of appropriate clinical interventions. As such virus isolation is

no longer a method of choice for routine diagnostics, although it remains an important approach in research.

3.2. Serology:

The principle of serological diagnostic test is to detect the presence of pathogen specific antibodies, often IgM/IgG in the tested specimens, including serum and CSF samples. A seroconversion is often required to establish the diagnosis. This method may be useful for the diagnosis of viruses such as flaviviruses (JEV, DENV and WNV) because the presence of these viruses in CSF is often transient (147,151). However, serological tests might fail to diagnose an infection due to delay in antibody response after symptom onset. Additional limitation is the high cross-reactivity usually occurring between closely related viruses, especially between JEV and DENV (152).

3.3. Nucleic acid based detection methods:

Molecular methods based on the detection of viral nucleic acids encompass viral specific polymerase chain reaction (PCR), pan viral family PCR and pan micro-array. These methods represent an advance for the diagnosis of viral infection because of its short turnaround, high sensitivity and specificity. A PCR assay procedure might take about 2-6 hours to complete (153), and the method can detect the presence of the pathogen genomes in the tested samples as low as 1-10 copies per PCR reaction. Because of these advantages, molecular methods, especially PCR are widely used of the diagnosis of viral infections (126,154–158).

In clinical settings, PCR methods are routinely used for the diagnosis of a wide range of viral pathogens such as HSV, measles, mumps virus, EVs, rabies viruses and flaviviruses (155,158). PCR is considered as gold standard for the diagnosis of HSV encephalitis. The reported sensitivity of CSF HSV PCR was around 98% (159), compared to the sensitivity of 50% of HSV

culture (160). However, CSF PCR is not useful for flavivirus detection at symptom onset, whereas the IgM ELISA is the test of choice (151).

A wide range of different PCR assays are often required to allow for the detection of a broad range of common viral pathogens in patient samples (26). Occasionally, multiplex PCR, which allow for simultaneous detection of multiple pathogens in a single test can be used, although the sensitivity can be compromised. Pan viral microarrays could detect several hundred viral species in a single test. It is however more relevant for research, and is often not available in resource constrained settings.

4. Sequencing technologies:

After being developed by Sanger in 1977 (161), Sanger sequencing technology quickly became the method of choice for sequencing because of its high throughput at the time and the requirement of less toxic chemicals (162). To this end, Sanger sequencing remains a reference sequencing method. However, due to the increasing demand of sequencing a large number of human genomes and other organisms, several next generation sequencing (NGS) technologies have been developed over the last two decades. Figure 1.3 shows the timeline of novel sequencing technologies that have been introduced over the last decades. The general features of NGS technologies and their applications in the field of virus detection and discovery are summarized in Table 1.6. Roche shut down 454 pyrosequencing in 2013; so the technique will not be discussed further in this thesis. I will be focusing on sequencing technologies developed by Illumina, Ion Torrent, Pacific Biosciences and Oxford Nanopore because these are common sequencing technologies used in the field of infectious diseases research and diagnostics, especially for pathogen detection and discovery.

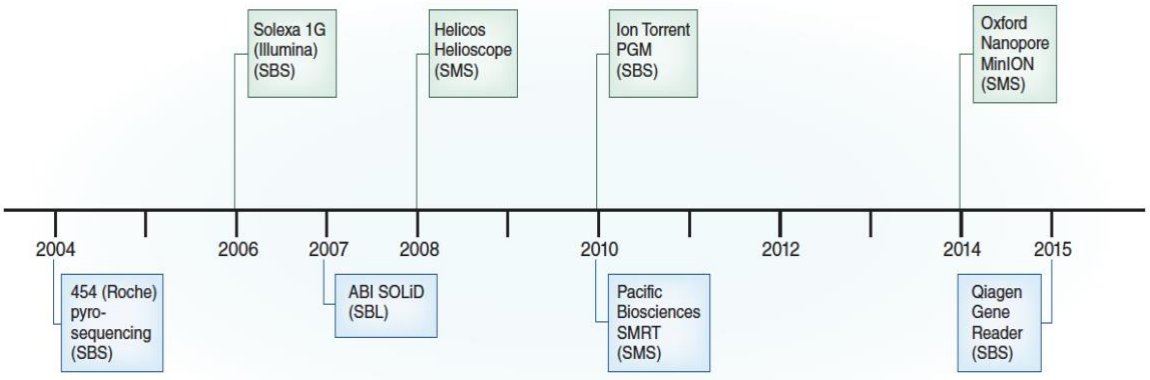


Figure 1.3 Timeline of introduction of next-generation sequencing technologies during past decades. SBS, sequencing by synthesis; SMS, single-molecule sequencing; SBL, sequencing by ligation. The figure was adapted from "DNA sequencing technologies : 2006 – 2016" by Mardis, 2017 (161).

Table 1.6 Summary of general features and applications of NGS technologies. The table was adapted from "DNA sequencing technologies: 2006 – 2016" by Mardis, 2017 (161)

NGS technologies	Sequencing chemistry	Output	Read length	Error rates	Applications
454/Roche	Sequencing by synthesis	700 MB	400 bp	High	Bacterial and viral genomes, multiplex-PCR products, validation of point mutations, targeted somatic-mutation detection
Illumina	Sequencing by synthesis	1.2– 330 Gb	150–300 bp	Low	Small and large genomes (humans, mouse, plants, bacteria, viruses etc) and genome-wide association studies, RNA-seq, hybrid capture or multiplex-PCR products, metagenomics, somatic-mutation detection, forensics, noninvasive prenatal testing
ABI Solid	Sequencing by ligation	2.5-9GB	50-75bp	Low	Complex genomes (human, mouse, plants) and genome-wide NGS applications, RNA-seq, hybrid capture or multiplex-PCR products, somatic-mutation detection
Ion Torrent	Sequencing by synthesis	10–1,000 MB	200–400 bp	High	Multiplex-PCR products, microbiology and infectious diseases, somatic-mutation detection, validation of point mutations
PacBio sequencing	Single-molecule sequencing	70 to 140 MB	10-50kbp	High	Complex genomes (human, mouse and plants), microbiology and infectious-disease genomes, transcript-fusion detection, methylation detection
Oxford Nanopore	Single-molecule sequencing	10-300 GB	up to 5.4 kbs	High	Pathogen surveillance, targeted mutation preparation detection, metagenomics, bacterial and viral genomes
Qiagen GeneReader	Sequencing by synthesis		107bp		Targeted mutation detection, liquid biopsy in cancer

4.1. Illumina:

Next generation sequencing (NGS) is deep, high-throughput, in-parallel DNA sequencing. It provides a massively parallel analysis with extremely high-throughput from multiple samples. Illumina technologies are the most commonly used sequencing system worldwide (163). Illumina purchased the Solexa Genome Analyzer in 2006 and commercialized it in 2007 (163). Illumina sequencing adopted the technology of reversible-termination sequencing by synthesis (SBS) with fluorescently labeled nucleotides. Illumina technology uses flow cells consisting of optically transparent slides with individual lanes. Small oligonucleotide anchors are immobilized on surface of each lane. The target template DNA is fragmented, phosphorylated at 5' end and adenylated at 3' end. Adaptors are ligated to DNA fragment. The adaptor-ligated oligonucleotides are complementary and attached to flow cell anchors. Cluster generation is featured by bridge amplification. DNA fragment flips over and forms a bridge by hybridizing to an adjacent and complementary anchor. Bound libraries are then extended by polymerase. On denaturation, double-strand molecules are separated. The original template is washed away. The newly synthesized strand is covalently attached to the flow cell surface. After multiple amplification cycles, a single DNA template is amplified as a clonally cluster consist of thousands of clonal molecules. Millions of clusters of different template molecules can be generated per flow cell (Figure 1.4).

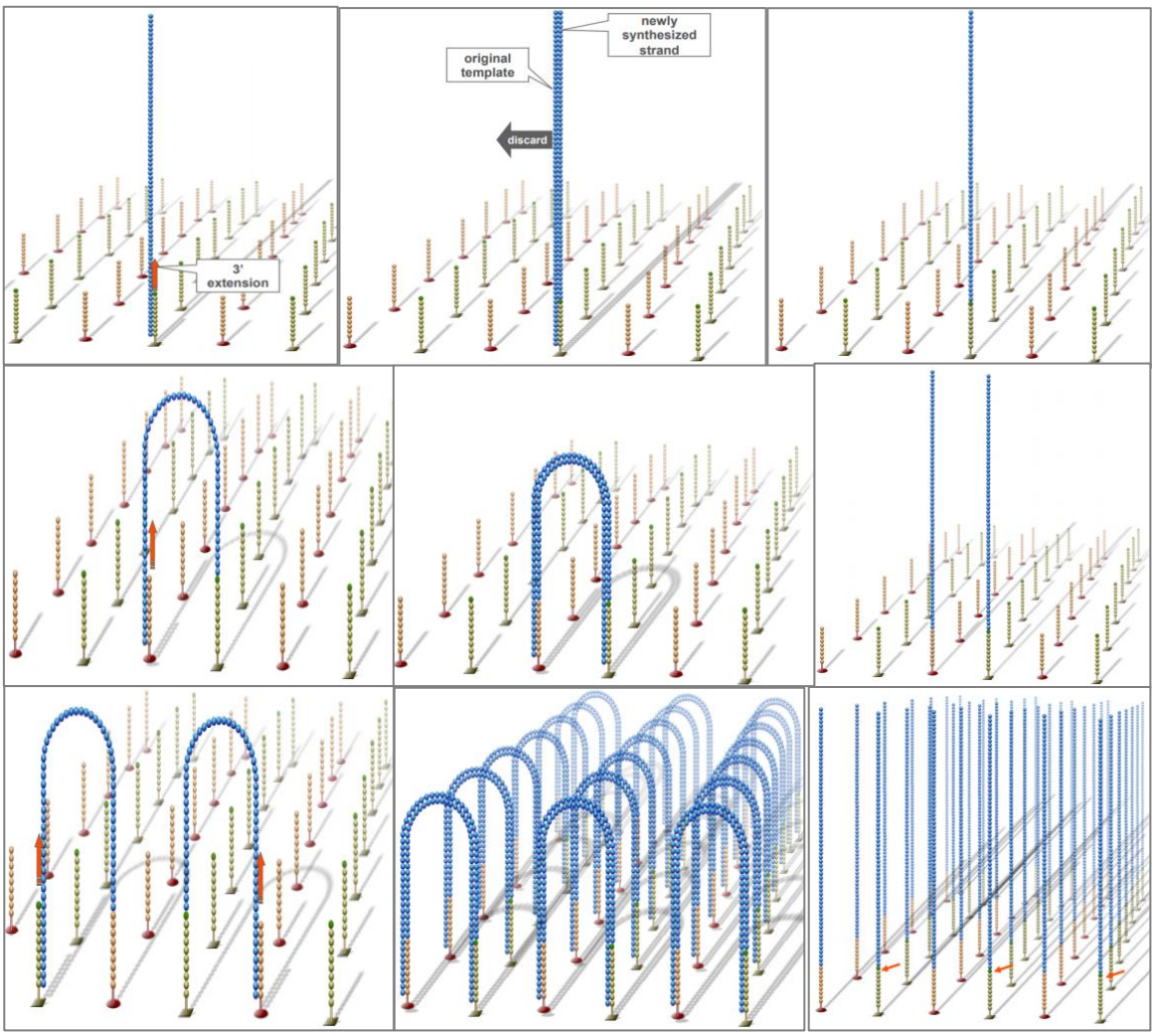


Figure 1.4 Illustration of cluster generation procedure of Illumina sequencing. Figure was adapted from" Explore Illumina sequencing technology: Massively parallel sequencing with optimized SBS chemistry" in https://sapac.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html (164).

Four fluorescently labeled nucleotides are used to sequence millions of clusters on the flow cell surface in parallel. In each growing chain, a single labeled dideoxynucleoside triphosphate (ddNTP) is added in each cycle. Due to the incorporation of the labeled nucleotide, DNA polymerization terminates, and the fluorescent dye is imaged to identify the incorporation. Then the fluorophore and terminator are enzymatically cleaved to allow for the incorporation of the next nucleotide (Figure 1.5) (162).

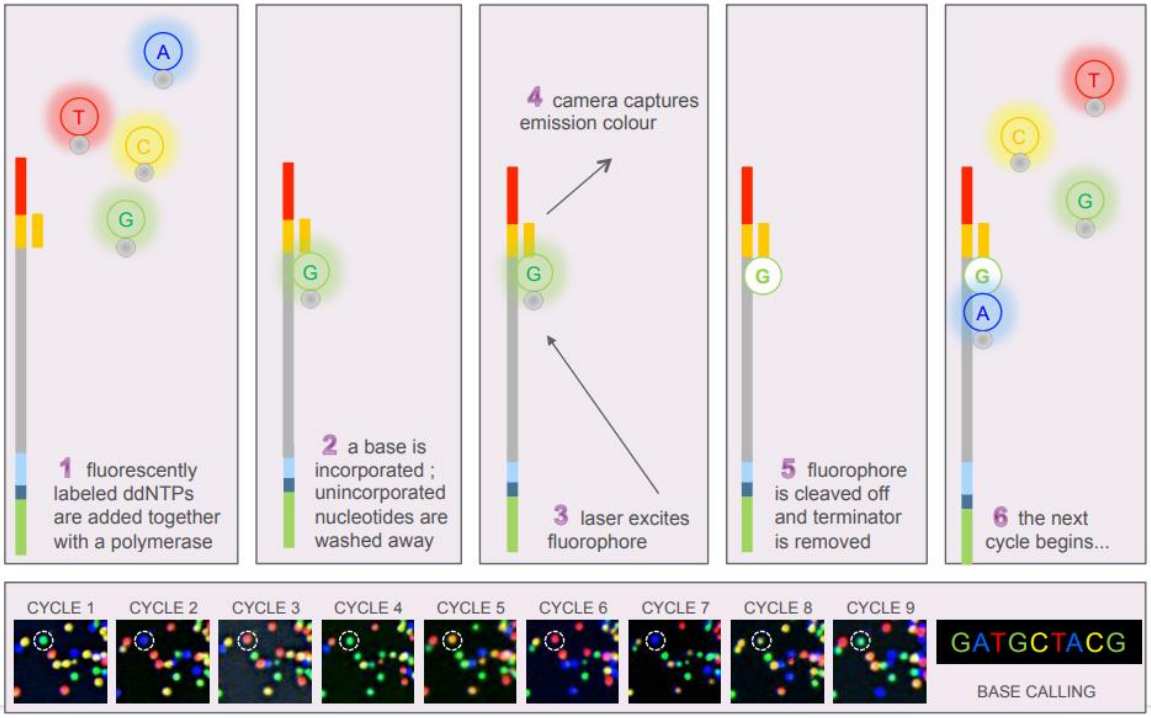


Figure 1.5 Illustration of sequencing by synthesis procedure of Illumina sequencing. Figure was adapted from" Explore Illumina sequencing technology: Massively parallel sequencing with optimized SBS chemistry" in https://sapac.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html (164).

Illumina provides five benchtop sequencer platforms (iSeq 100, MiniSeq, MiSeq, NextSeq 550 and NextSeq 1000&2000) with a wide range of output (1.2Gb – 330Gb) with short-length reads (150-300bp) (Table 1.7). Generating high throughput with low errors (accuracy of over 99%) (165), Illumina sequencing has thus becomes the most commonly used technology (including in the field of virus discovery) (166–170). MiSeq, a compact laboratory sequencer, has an output of 0.3-15Gb and can provide sequencing results within 2-3 days. It is currently widely used (165,171–177).

Table 1.7 Summary of features and application of Illumina benchtop sequencers.

Sequencer platforms	Maximum Output (Gb)	Maximum read length (bp)	Run time (hours)	Applications
 iSeq 100	1.2	2 × 150	9.5–19	• Small Whole-Genome Sequencing • Targeted Gene Sequencing • Targeted Gene Expression Profiling • miRNA & Small RNA Analysis
 MiniSeq	7.5	2 × 150	4–24	• Small Whole-Genome Sequencing • Targeted Gene Sequencing • Targeted Gene Expression Profiling miRNA & Small RNA Analysis • 16S Metagenomic Sequencing
 MiSeq	15	2 ×300	4–55	• Small Whole-Genome Sequencing • Targeted Gene Sequencing • Targeted Gene Expression Profiling miRNA & Small RNA Analysis • 16S Metagenomic Sequencing • DNA-Protein Interaction Analysis
 NextSeq 550	120	2 × 150	12–30	• Small Whole-Genome Sequencing • Targeted Gene Sequencing • Targeted Gene Expression Profiling miRNA & Small RNA Analysis • 16S Metagenomic Sequencing • DNA-Protein Interaction Analysis • Exome & Large Panel Sequencing • Single-Cell Profiling • Transcriptome Sequencing • Methylation Sequencing • Metagenomic Profiling • Cell-Free Sequencing & Liquid Biopsy Analysis
 NextSeq1000&2000	330	2 × 150	11-48	• Small Whole-Genome Sequencing • Targeted Gene Sequencing • Targeted Gene Expression Profiling miRNA & Small RNA Analysis • 16S Metagenomic Sequencing • DNA-Protein Interaction Analysis • Exome & Large Panel Sequencing • Single-Cell Profiling • Transcriptome Sequencing • Methylation Sequencing • Metagenomic Profiling • Cell-Free Sequencing & Liquid Biopsy Analysis

4.2. Ion Torrent:

Ion Torrent sequencing technology was released by Life Technologies/Thermo Fisher Scientific in 2010. The principle of this technology is detection of hydrogen ion releasing during incorporation of new nucleotides into the growing DNA template. Particularly, adapter-ligated DNA templates are clonally amplified by emulsion-PCR on the surfaces of beads. These beads are distributed into micro wells where a sequencing-by-synthesis reaction occurs. An ion sensor is used to detect signal from protons releasing during nucleotide incorporation (Figure 1.6) (178). Despite of generating low output, Ion Torrent offer low cost and short run time. Therefore, it is a reasonable choice in some cases, such as detection of targeted virus in clinical samples (179) or viral genome sequencing (180).

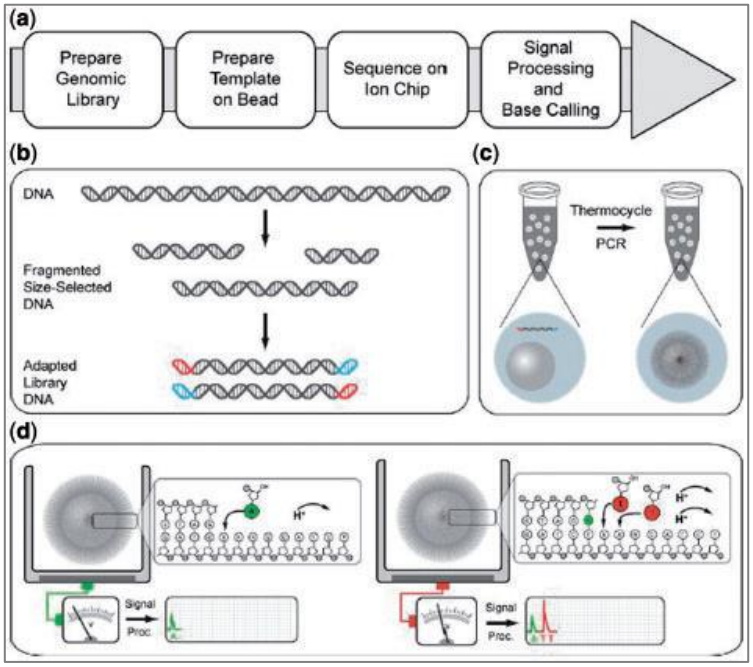


Figure 1.6 Illustration of overall workflow of Ion torrent sequencing. (a) library preparation, (b) clonal amplification of adapter-ligated template on a bead, (c) binding of sequence on chip, sequencing primers and DNA polymerase to the beads by pipetting into wells on chip, (d) sequencing-by-synthesis reaction. At each flow, the electrical signal at each well is measured, indicating the number of incorporations. The figure was adapted from "Using state machines to model the Ion Torrent sequencing process and to improve read error rates" by Golan, 2013 (178).

4.3. PacBio sequencing:

The technology was introduced by Pacific Biosciences in 2010, which is a single molecule real time sequencing platform for generating long reads without bias of clonal amplification. PacBio sequencing uses Zero Mode Waveguide (ZMW), a nano-hole made in a 100 nm metal film on a glass surface. In brief, a single DNA polymerase molecule is anchored at the bottom of the ZMW. Four types of nucleotides labeled with different colored fluorophore are flooded above the ZMWs. Labeled nucleotides travel down into the ZMW within microseconds. After reaching the DNA polymerase, they diffuse back up and exit the hole. During polymerization, fluorescent tag of nucleotides is cleaved off and diffuses out of the observation area of the ZMW and detected by detector in real time (Figure 1.7). Single molecule sequencing with long read length is ideal for viral metagenomics (181). However, PacBio sequencing has several limitations including low throughput with high error rate and high cost per sequenced base.

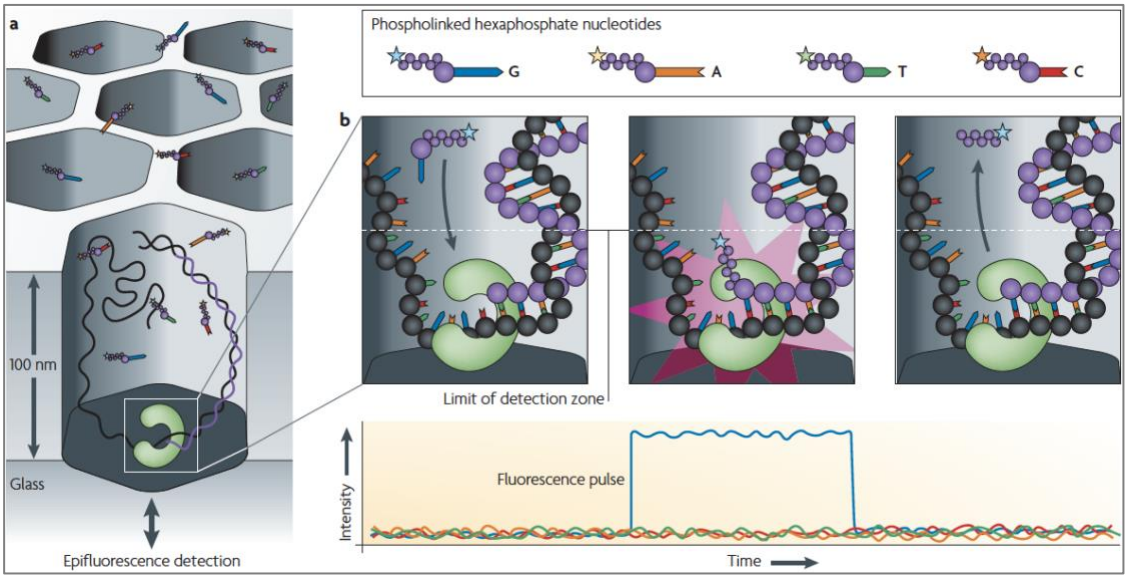


Figure 1.7 Illustration of Pacific Biosciences' four-color real-time sequencing method. Figure was adapted from "Sequencing technologies-the next generation" by Metzker, 2009 (182).

4.4. Oxford Nanopore sequencing:

Oxford Nanopore sequencing or third generation sequencing was developed with the aim to generate long reads with low cost and simple sample preparation procedures. MinION, a portable sequencer of Oxford Nanopore Technologies (Figure 1.8) was first released in 2014. It is the first handheld device for DNA and RNA sequencing that attaches directly to a laptop/computer via a USB port.

Figure 1.8 MinION flow cell

Nanopore sequencing works on the principle of measuring minute changes in electric current across the nanopore immersed in a conducting fluid with voltage applied that are induced when a moving biological molecule (DNA/RNA) passes through it. Nanopore is a small hole with internal diameter of 1nm that is made up of certain transmenbrance cellular protein (162). The sequence of the passing DNA is identified by detecting changes of current generated specific for passed bases. The advantages of Nanopore are that it is the first device could deliver real-time sequencing of single molecule, and able to generate very long DNA molecules (up to 5.4 kbs genome) in a single read. The MinION platform was successfully used to sequence ebola virus genomes during a recent outbreak in the Democratic Republic of Congo (183). MinION was also successfully applied to track the transmission of ZIKV during the outbreak in South

America in 2015 (184). Most recently, MinION was used to generate complete genome sequence of SARS-CoV-2 (185).

High-throughput sequencing benchtop platforms were released including GridION and PromethION. These platform incorporate multi flow cells in one therefore could generate higher yield with comparable prices (186). However, current limitation of MinION is its high error rate of around 10% compared to the low error rate of Illumina sequencing technologies (185).

5. Metagenomic next generation sequencing for virus detection and discovery:

Metagenomic next-generation sequencing (mNGS) is a generic procedure that can sequence all type of genetic materials, including those derived from the pathogens present in a sample under investigation (187). mNGS overcomes the limitations of conventional diagnostic methods such as PCR and culture, because it does not require prior knowledge/assumptions about the targeted causative agents present the tested samples (188,189). In principle, in a single assay mNGS could detect all viral agents (including known and previously unknown viruses) in any clinical sample types (188). This strategy could detect wide range of virus pathogens in the tested samples and offer a new opportunity for virus detection and discovery. From a literature review, around 35 novel viruses or new viral variants were discovered by mNGS assays from 2008 to 2019. The number of new discovered viruses by mNGS has gradually increased over the last 10 years with peak at 2015 (Figure 1.9).

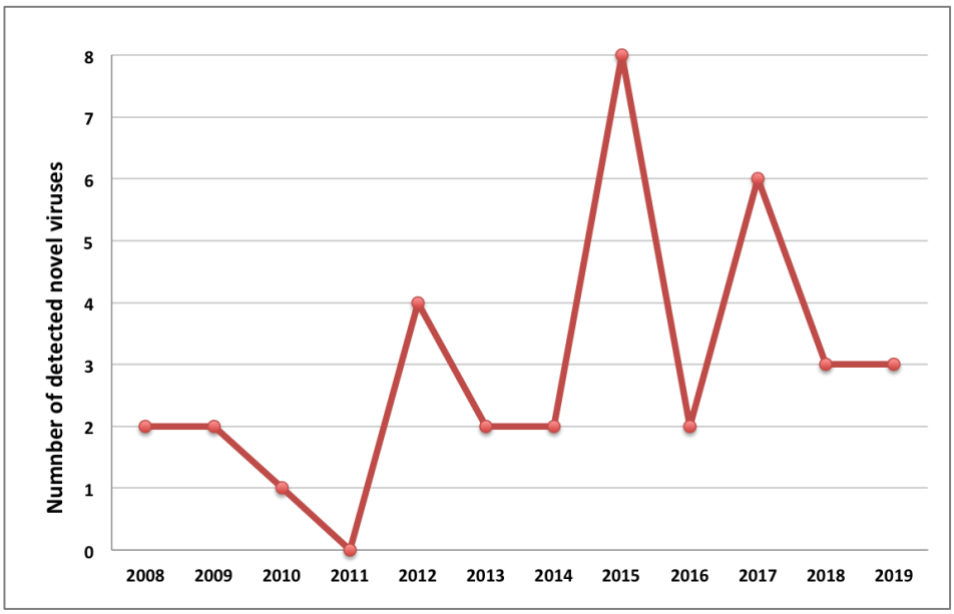


Figure 1.9 Number of novel viruses recently detected by mNGS assay during 2008-2019

However, mNGS based detection of viral infections remains a challenge because viruses are naturally small in size, genetically highly diverse, and have a low abundance as compared to the host DNA background in clinical samples. Therefore, various sample pretreatment approaches have been developed to increase the sensitivity of mNGS. The commonly used approaches are descried herein.

5.1. Pre-extraction viral enrichment methods:

Pre-extraction viral enrichment methods are based on the physical properties of virions, including the size of virus particles, their density, and the presence or absence of viral capsid covering the virus genome (189). Accordingly, three complementary approaches for viral nucleic acid purification and enrichment have been used: filtration, density gradient centrifugation and enzymatic removal of non-capsid protected nucleic acids.

5.1.1. Filtration:

Filtration is a common method used for purifying viral particles from environments or clinical specimens prior sequencing. The size of most animal viruses ranges from 200 to 300 nm in diameter. Accordingly, the common pore sizes of filters are often from 0.2 to 0.45 micrometer (174,190–192).

5.1.2. Ultra centrifugation:

The compact nature of viral particles allows for the use of density gradient centrifugation method to purify viral nucleic acids. The differences in density of enveloped and non-enveloped viruses require the collection and subsequent analysis of several gradient fractions. Moreover, the presence of contaminating nucleic acids in the density gradient solution can lead to subsequent contamination in target sample (189). Therefore, this method is not widely applied for the analysis of clinical samples.

5.1.3. Nuclease treatment:

Nuclease treatment prior to viral nucleic acid extraction is an efficient approach to digest naked host DNA present in clinical samples. The incorporation of DNase I treatment for virus discovery approach was firstly reported by Allander et al. (193). It was a key step for removal of host DNA in serum prior amplification of viral nucleic acids. DNase I treatment removes naked DNA by its exonuclease digestion activity, while viral DNA is protected within viral capsids. Similarly, RNase treatment is used for removal of exogenous RNA in clinical samples (194,195).

5.2. Enrichment of viral nucleic acids prior to sequencing:

Viral genomes are much smaller than the human genome. Therefore, the proportion of viral reads relative to host derived reads in mNGS data would be extremely low. Selective amplification of viral nucleic acids prior to sequencing is critical to enhance the sensitivity of mNGS, thereby increasing the chance of detecting a virus in the tested samples (188).

Commonly used approaches include sequence-independent single primer amplification, random PCR and PhiX29.

5.2.1. Sequence-independent single primer amplification (SISPA):

SISPA utilized restriction enzymes to first digest the targeted DNA, followed by the ligation of adaptors complementary to the overhanging ends of the target DNA. For RNA virus, a cDNA synthesis step followed by double strand DNA (dsDNA) synthesis is required prior to the restriction enzyme digestion step. Next the ligated products are amplified using adaptor specific primer. Before NGS becomes available, after the amplification step, the amplified products are subcloned, then plasmid inserts are sequenced (193). The method was successfully used to characterized a Norwalk virus from faeces (196) and a human astrovirus from culture supernatants (197), new parvoviruses in human plasma (198). Nowadays, SISPA products are subjected to NGS without the cloning step.

5.2.2. Random PCR:

Random PCR (rPCR) utilizes a primer consisting of a random or viral specific hexamer at the 3' end and a unique second primer sequence at the 5' end (188). The hexamer part is used for the synthesis of the cDNA and ds DNA. During this step, the second part of the primer sequence (5' end sequence) is thus introduced into the resulting double-stranded sequences. It will then serve as the binding site of the PCR primers during the amplification step. Subsequently the resulting PCR products, including those derived from viral nucleic acids are sequenced.

rPCR method theoretically could amplify all genetic materials present in the tested sample. rPCR based approaches have been used to characterize a new parvovirus and numerous viral agents in respiratory secretions (199), a new gyrovirus in children's faeces (200), and a new bunyavirus in serum samples (201).

5.2.3. PhiX29:

PhiX29 DNA polymerase based amplification method is based on the ability of bacteriophage PhiX29 DNA polymerase to generate multiple displacement amplification reactions. The DNA polymerase is primed with a set of modified random hexamer oliogonucleotides, which is resistant to the 3'-5' exonuclease activity of PhiX29. This method successfully amplifying circular DNA anellovirus (202,203).

5.3. Bioinformatics:

Advanced bioinformatic methods are required to detect virus sequences from NGS data and to reconstruct the sequenced viral genomes. Many published computational workflows with a range of computer tools for taxonomic classification have been developed to analyse metagenomic data. They could be web-services with a graphical user-interface working fast on any PC such as Taxonomer (204), Kraken (205) or IDseq (206). Or they can be command line pipelines such as SURPI (207) or in-house viral bioinformatic pipeline (190). The processing time per sample ranges from minutes to several days (208). The web-services are friendly to use, but they often require uploading large output NGS data to a distant server.

In general, a metagenomic workflow basically consists of five steps: pre-processing, filtering out DNA sequences from the respective hosts, assembly, searching for viral reads and post-processing (208). The pre-processing step may include: removing adapter sequences, trimming and low quality reads, removing low complexity reads, short reads, duplicating reads, matching paired-end reads and unresolved nucleotides. The second step is to filter out non-viral reads. This step helps to reduce false positive results and prevent assembly of chimearic virus-host sequences. The third step is to assemble the short mNGS into long contigs. Consensuses may be generated by mapping individual reads to these obtained contigs. This is also regarded as de novo assembly. This step allows for remove errors from individual reads and reduce the amount of data for further analysis. The fourth step is to conduct database searching whereby contigs

are mapped to a reference database. The searching approach is based on the alignment the reads or contigs to a reference database (199). The reference database can be retrieved from the NCBI GenBank, RefSeq or BLAST nucleotide and non-redundant protein databases. Searching at protein level allow for detecting more remote homology that enhance detection of highly divergent viruses (188). Finally, the post-processing step is to classify the sequencing reads. This is the process of finding the most likely or best-supported taxonomic assignment among reference database. This step can use phylogenetic or other computational methods.

5.4. Application of viral mNGS in sepsis:

There have been several published papers reporting the use of mNGS for the diagnosis for bloodstream infection. However, to the best of my knowledge there have been few studies describing the use of mNGS to study the etiology, especially bacterial causes in patients presenting with sepsis or sepsis shock (209–212). mNGS methods successfully detected HSV-1 sequences in Chinese fatal patients with sepsis of unknown origin (210) and 15 viral pathogens in 14 blood samples of ICU patients with suspected sepsis (212).

5.5. Application of viral mNGS in CNS infection:

A recent literature review showed that there has been an exponential increase in the number of studies utilizing mNGS for the detection of CNS pathogens (Figure 1.10) (175).

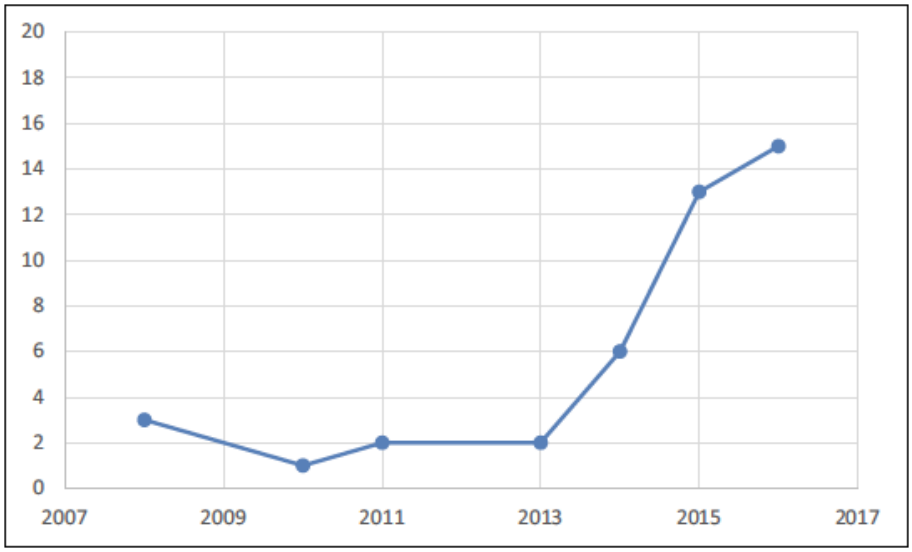


Figure 1.10 Temporal trends in the Publication of Encephalitis Cases involving Next-Generation Sequencing in the last Decade. (Brown et al, 2018)

Although the clinical studies were conducted in many countries worldwide, the laboratories performing the analysis were mainly based in developed countries such as the US and in Europe (175).

Often mNGS is used when the diagnostic work-up based on conventional assays failed to identify a pathogen (173,177,213–215). The diagnostic yield of mNGS might be sample dependent. Indeed, previous studies showed brain biopsies gave a higher yield than CSF did (175,216). Due to the uncertainty about the specificity of mNGS, confirmatory testing of mNGS results is needed. For this purpose, viral-specific PCR assays are commonly used. In contrast to cell culture, serological assays and immunohistochemistry methods are rarely used.

6. Global emerging infectious disease hotspots:

Worldwide, there are over 320,000 mammalian viruses remaining uncharacterized (217). Of note, Asia including Vietnam is one of the major hotspots for the emergence of novel pathogens as illustrated by the emergence of Nipah in 1998, SARS-CoV-1 in 2003, avian influenza A virus subtype H5N1 in 2004, enterovirus A71 and influenza A virus subtype H7N9 in 2013. Most recently, SARS-CoV-2 was first discovered in a patient suffering from community-acquired pneumonia of unknown origin in China in late 2019 (218). SARS-CoV-2 has now spread globally and is responsible for the ongoing COVID-19 pandemic. The high density of the human population and the fact that humans are living in a close proximity to domestic and wildlife animals in Asia are considered risk factors of disease emergence.

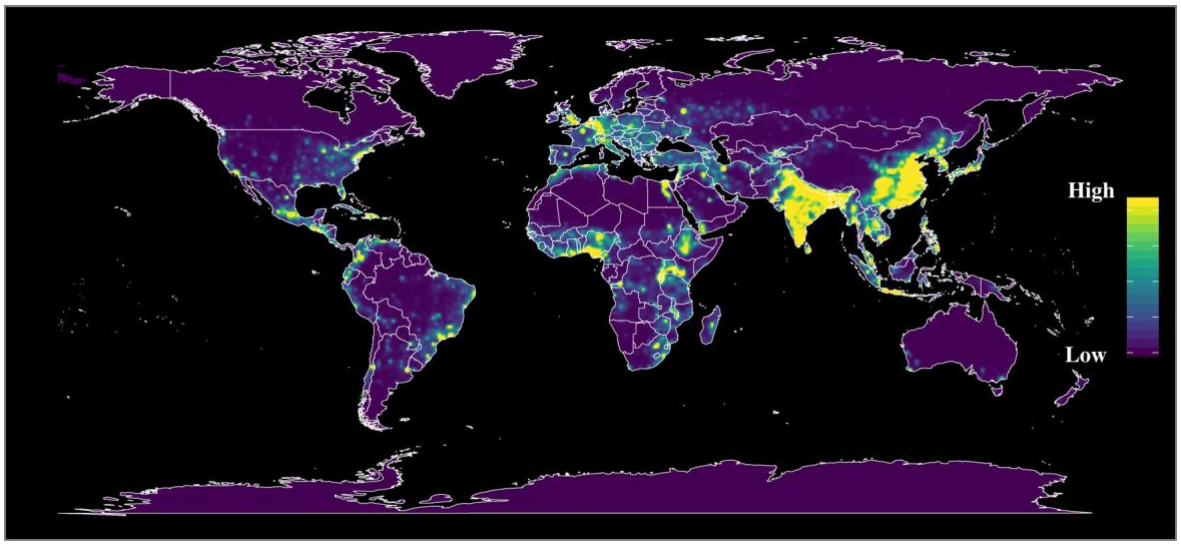


Figure 1.11 Heat maps of predicted relative risk distribution of zoonotic emerging infectious disease events. Figure was adapted from "Global hotspots and correlates of emerging zoonotic diseases" by Allen et al, 2017 (219).

7. Aims:

Collectively, studies to date (including those from Vietnam) have demonstrated a wide range of causative agents (especially viruses) can cause CNS infections and sepsis. Yet, despite the use of advanced molecular diagnostic assays (PCR and metagenomics), it remains a challenge to identify a viral culprit in patients with CNS infections. Few studies however from Vietnam and Asia reported the use of metagenomics to actively study the etiology in patients with CNS infections and sepsis. In this region of the world, novel CNS infections pathogens (EV-A71, Nipah, hendra and Zika viruses) are however likely to emerge. Therefore, improving our knowledge about the causes of sepsis and CNS infections, and active surveillance for novel pathogens in Southeast Asia are of clinical and public health significance. I hypothesize that next-generation sequencing based viral metagenomics will identify known or unknown viruses in undiagnosed patients with CNS infections and community-acquired sepsis in Southeast Asia. Therefore, within my PhD research program, I aimed to:

1. Develop a sensitive viral metagenomic pipeline for sequence-independent detection of a broad range of viral pathogens in clinical samples
2. Explore viral content in patients with sepsis of unknown cause across Southeast Asia
3. Explore viral content in CSF from patients with acute CNS infections of unknown cause sampled from provincial hospitals throughout Vietnam
4. (If relevant), demonstrate proof of causation of recently described viruses/novel virus(es) discovered by metagenomic analysis
5. Explore the utility potential of metagenomics for the diagnosis of central nervous system infections

Chapter 2: Development of a sensitive viral metagenomic pipeline for sequence-independent detection of a broad range of viral pathogens in clinical samples

1. Background:

Designed by Endoh et al in 2005 (220), non-ribosomal primer consists of a set of 96 hexanucleotides, which are specific for amplification of viral sequences. Using this primer set, I have recently developed a non-ribosomal random amplification and next-generation sequencing based assay for sensitive detection and direct whole-genome sequencing of hand, foot and mouth disease pathogens from throat and rectal swabs (221). For my PhD research, I further optimized this method for sensitive sequence-independent detection of a broad range of viral pathogens in patient samples.

Clinical samples may contain low pathogen load but high concentration of contaminating host and bacterial nucleic acids. To enrich for viral nucleic acid and to reduce the unwanted background, sample pretreatment steps such as centrifugation and nuclease (DNase and RNase) digestion are often employed as part of viral metagenomic pipelines. However, few studies have assessed the efficiency of these pretreatment strategies. To identify to the optimal approach for viral metagenomic analysis, herein these sample pretreatment steps were compared. For the purpose of assay comparison, the number of viruses detected and the level of genome coverage were taken into account.

2. Methods and materials:

2.1. Patient samples and positive controls:

A biological reagent containing 25 different DNA and RNA viral pathogens prepared for viral metagenomic pipeline evaluation by National Institute for Biological Reagents and Control, UK (190) was used. Additionally, sera, CSF and zika virus culture spiked in human sera were also

use. The PCR positive clinical samples included a pooled serum sample derive from sera that were positve for either hepatitis A/B/C virus (222), and a total of six CSF samples that were PCR positive for dengue virus (n=3), herpes simplex virus (n=2) and mumps virus (n=1) (223). The zika virus (strain # MR766) was obtained from the European Virus Archive.

2.2. Methods:

2.2.1. Sample pretreatments and nucleic acid isolation:

A combination of i) sample pre-centrifugation and/or ii) RNase/DNase digestion was taken into account for assay development. Consequently, four different combinations of sample pretreatments were selected for comparison, including 1) simultaneous DNase and RNase treatment of the original sample without pre-centrifugation, 2) DNase treatment of the original sample without RNase treatment and pre-centrifugation, 3) sample pre-centrifugation followed by DNase treatment of the resulting supernatants, and 4) sample pre-centrifugation followed by simultaneous DNase and RNase treatment of the resulting supernatants. Specifically, prior to nucleic acid (NA) isolation 110µl of clinical sample was centrifuged for 10 minutes at 13,000rpm and then 100µl supernatant was collected. For both collected supernatant and non-centrifuged samples, 100µl was treated with 2U/ul of turbo DNase and with/without 0.4U/ul RNase I (Ambion, Life Technology, Carlsbad, 130 CA, US) at 37 °C for 30 min. The treated sample was then preceded to a viral NA isolation step using QIAamp viral RNA kit (QIAgen GmbH, Hilden, Germany), and finally recovered in 50ul of the elution buffer provided with the extraction kit.

2.2.2. Double stranded DNA synthesis and Sequencing:

Double stranded DNA was then synthesized from the isolated viral NA using a set of 96 non-ribosomal random primer (Appendix 1) (220), followed by PCR amplification to enrich for viral NA prior to sequencing (224,225). Finally the amplified products were sequenced on an Illumina MiSeq platform (Illumina, San Diego, CA, US) available at OUCRU.

In details, 10µl of extracted RNA was firstly mixed with 2µl of non-ribosomal random primer mixture and 1µl of dNTPs (10mM each) (Roche Diagnostics GmbH, Mannheim, Germany). The mixture was incubated at 65ºC for 5 min, and was then immediately chilled on ice for 1 min. Secondly, 7µl of a reaction mix containing 200U of Super Script III reverse transcriptase (Invitrogen, Carlsbad, CA, US), 40U of RNase OUT (Invitrogen), 0.1M DTT (Invitrogen) and 5X first strand buffer (Invitrogen) was added into the first reaction mixture. The reaction was then continued at 25ºC for 10 min, 37ºC for 1 min and 94ºC for 2 min, and immediately chilled on ice for 2 min. Next, 5U of exo-Klenow fragment (Ambion) and 10U of Ribonuclease H (Ambion) were then added into the reaction mixture. The mixture was subjected to a thermal condition consisting of 25ºC for 5 min, 37ºC for 1h and 94ºC for 2 min. This exo-Klenow fragment associated step was repeated once more time with the omission of the last incubation step on ice after the second thermal cycle.

Finally, 5µl of the resulting dsDNA was pre-amplified using FR20RV primer (5'-GCCGGAGCTCTGCAGATATC-3'). Random amplification (rPCR) was carried out in a total reaction volume of 50µl consisting of 3µl of dsDNA, 2µl of primer FR20RV at a final concentration of 40nM and 45µl of Platinum PCR supermix (Invitrogen). The thermal cycling condition consisted of 94ºC for 2 min and followed by 40 cycles of 94ºC for 30s, 55ºC for 30s and 72ºC for 3min and 1 cycle of 72ºC for 2min. List of reagents and thermal conditions used for these procedures are summarized in Table 2.1.

Table 2.1 List of reagents and thermal conditions of pretreatment, dsDNA synthesis and random amplification procedures.

Procedures	Reagents	Concentration	Used volume (per reaction)	Thermal cycling condition
Pretreatment	Turbo DNase	2U/µl	10µl	37 °C for 30 min
	RNase I	100U/µl	1µl	
	DNase buffer	10X	12µl	
Double strand DNA synthesis	Non-ribosomal primer mixture	1µM	2µl	85ºC for 2 min
	dNTPs	10µM	1µl	
	Super Script III reverse transcriptase	200U/µl	1µl	25ºC for 10 min, 37ºC for 1 min and 94ºC for 2 min
	RNase OUT	40U/µl	1µl	
	DTT	0.1M	1µl	
	First strand buffer	5X	4µl	
	exo-Klenow fragment	5U/µl	0.5µl	25ºC for 5 min, 37ºC for 1h and 94ºC for 2 min
	Ribonuclease H	10U/µl	0.5µl	25ºC for 5 min, 37ºC for 1h and 75ºC for 10 min
Random amplification	FR20RV primer	10µM	1µl	94ºC for 2 min, 40 cycles of 94ºC for 30s, 55ºC for 30s and 72ºC for 3min and 1 cycle of 72ºC for 2min
	Platinum PCR supermix		22µl	

The obtained random PCR product was then purified with use of QIAquick PCR purification kit (QIAgen GmbH, Hilden, Germany) and quantified by Quibit dsDNA HS kit (Invitrogen). Finally, 1ng of purified product was subjected to library preparation using Nextera XT sample preparation kit (Illumina, San Diego, CA, US). Prior to sequencing, the quantity of the prepared library was measured by using KAPA Library Quant Kit (Kapa Biosystems, Wilmington, MA, US). The prepared library was sequenced by using MiSeq reagent kit v3 (600 cycles) (Illumina, San Diego, CA, US) in a MiSeq platform (Illumina, San Diego, CA, US). For each run, samples were multiplexed and differentiated by double indexes using Nextera XT Index Kit (Illumina, San Diego, CA, US).

2.2.3. Sequence analysis and statistical analysis:

The obtained MiSeq data was analysed using publically available bioinformatic pipelines, including SURPI and Taxonomer (204,207,226) to identify the presence of viral sequences in the tested specimens. A reference-based mapping approach (Geneious 8.1.5) was then employed to assess the level of genome coverage of the sequenced viruses.

Chi-square test (Prism 9.0.0) was applied to measure the differences in the number of viral used reads generated by different sample-pretreatment approaches. A p value of less than 0.05 was considered statistical significance.

2.2.4. Contributions from others:

Clinical samples and data collection were carried out by participating clinicians and research staff of the respective collaborating hospitals. Although, I led the experiment design and conduct the most of the laboratory work, I also received support in conducting some of the PCR confirmatory experiments from Ms Le Nguyen Truc Nhu and Ms Nguyen Thi Thu Hong from Emerging Infections group, OUCRU.

3. Results:

3.1. Efficiency of viral metagenomic approaches in detecting a wide range of viruses:

Table 2.2 summarized the results of viral detection by 4 different sample-pretreatment approaches followed by random PCR amplification and Illumina MiSeq sequencing. In terms of viral detection, 16-18 out of 25 viruses included in the original biological reagent were detected by the viral metagenomic approaches under comparison, of which approaches #1 (simultaneous DNase and RNase treatment of the original sample without pre-centrifugation) and #2 (DNase treatment of the original sample without RNase treatment and pre-centrifugation) gave the best performances (Table 2.2 and 2.3). Indeed, approaches #1 and #2 generated a higher proportion and number of viral reads mapped to a reference genome than approaches #3 and #4 did ($p<0.0001$, Table 2.2).

In terms of genome coverage, of the 18 detected viruses, 5 complete or nearly complete genome sequences (Appendix 2) were generated by approach #1, while 1-3 complete or nearly complete genome sequences were generated by the other approaches (Table 2.2). Six out of 7 viruses that were not detected by my assay, were real-time PCR negative after pooling (190). My assay had

a comparable sensitivity with metagenomic pipelines developed by other groups elsewhere (190,227) (Table 2.3).

Collectively, based on the obtained result, the viral metagenomic approach utilizing a simultaneous DNase and RNase treatment step of the original sample without pre-centrifugation was the most sensitive approach and was therefore selected for additional evaluation on clinical samples.

Table 2.2 Number of total reads and viral reads generated by viral metagenomics using different samples pretreatment approaches

	#1 +DNase+RNase -Centrifugate	#2 +DNase-RNase -Centrifugate	#3 +DNase-RNase +Centrifugate	#4 +DNase+RNase +Centrifugate
Number of total reads	2,102,000	2,225,000	2,171,000	2,072,000
Number of reads mapped to a reference viral genome (%)	801774 (38)	884507 (39)	72533 (3.5)	62820 (2.8)
Number of viral species detected	18	18	16	16
Number of viruses detected with a complete genome obtained	5	3	1	1

Table 2.3 Summary of viral metagenomics detection results for viral mixture using different sample pretreatment approaches.

Species	Ct values (190)	Numbers of viral reads detected by my metagenomic approaches				Numbers of viral reads detected by viral metagenomic pipelines from other studies	
		#1 +DNase+RNase -Centrifugate	#2 +DNase-RNase -Centrifugate	#3 +DNase-RNase +Centrifugate	#4 +DNase+RNase +Centrifugate	US study (190)	Zurich study (227)
Adenovirus 2	29.71	998	800	140	244	260	299.1
Astrovirus	30.53	1388	761	1003	823	14	8040.6
Coxsackievirus B4	30.72	4765	6219	3946	7735	24	22,707
Human herpesvirus 1	30.59	38	0	0	0	11	81.4
Human herpesvirus 2	32.48	176	120	0	0	7	7
Human herpesvirus 3	29.02	68740*	54414*	19331	14516	330	84.2
Human herpesvirus 4	31.27	1246	662	425	509	34	104.7
Human herpesvirus 5	28.95	21000	14676	7517	3737	447	7028.5
Human Metapneunmovirus A	31.86	10938*	2041*	756	1364	26	0
Influenza A virus H1N1	32.02	251	363	0	0	2	0
Parainfluenzavirus 1	34.43	3794*	3129	3085	2215	44	9601.2
Parainfluenzavirus 2	33.87	25286*	477	60	6	253	33.1
Parainfluenzavirus 4	31.83	4508	2696	1929	1404	24	10,089
Parechovirus 3	29.35	637000*	772000*	1073*	1204*	3507	565.22
Respiratory syncytial virus A2	34.33	10	18	17	36	4	229.4
Rhinovirus A39	31.16	1475	598	340	1011	6	2238.4
Rotavirus A	24.9	20054	25391	32634	27668	2896	8.5
Sapovirus C12	33.37	107	98	275	252	14	62.8
Adenovirus 41	ND	0	0	0	0	4	13.9
Coronavirus 229E	36.48	0	0	0	0	0	0
Influenza A virus H3N2	ND	0	0	0	0	0	0
Influenza B virus	ND	0	0	0	0	0	0
Norovirus GI	ND	0	0	0	0	0	0
Norovirus GII	ND	0	0	0	0	0	0
Parainfluenzavirus 3	ND	0	44	2	96	0	0
Total number of detected viruses		18	18	16	16	19	17

Note: ND: not detected by real time PCR after pooling, *complete/nearly complete coding sequence were obtained.

3.2. Viral metagenomic assay performance on clinical samples:

The selected viral metagenomic assay (approach #1) was further evaluated on clinical samples (including CSF and serum) that were PCR positive, or zika virus culture spiked in human sera. Subsequently, the assay was able to detect the expected viral pathogens in the corresponding tested materials (Table 2.3).

Table 2.3 List of viral pathogens used for further evaluation of the viral metagenomic assay

Virus	Sample type	Viral type	Clinical presentation	Diagnostic Ct value	Number of viral reads	Percentages of genome coverage
Hepatitis A virus	Pooled serum*	(+) ssRNA	Hepatitis		3	NA
Hepatitis B virus		dsDNA	Hepatitis		3,256	100%
Hepatitis C virus		(+) ssRNA	Hepatitis		10	NA
Zika virus (10^-1)[#]	virus culture spiked in a human serum sample	(+) ssRNA	NA	20	252,732	99%
Zika virus (10^-3)[#]	virus culture spiked in a human serum sample	(+) ssRNA	NA	27	2,135	80%
Dengue virus 1[$]	Cerebrospinal fluid	(+) ssRNA	CNS infections	32	13,813	96%
Dengue virus 1[$]	Cerebrospinal fluid	(+) ssRNA	CNS infections	32	1,467	50.4%
Dengue virus 2	Cerebrospinal fluid	(+) ssRNA	CNS infections	39	4	NA
Herpes simplex virus	Cerebrospinal fluid	dsDNA	CNS infections	24	8	NA
Herpes simplex virus	Cerebrospinal fluid	dsDNA	CNS infections	27	2	NA
Mumps	Cerebrospinal fluid	(-) ssRNA	CNS infections	34	14	NA

Note: * derived from three different sera positive for one of these three viruses, [#] dilution ratio of viral culture in serum samples, NA: none applicable, [$] two different specimens.

4. Discussion and Conclusion:

In this chapter, I set out to develop a viral metagenomic pipeline for broad-range detection of viral pathogens in clinical samples. The obtained results showed that the metagenomic assay employing a DNase/RNase treatment step of the specimen without sample pre-centrifugation is a sensitive method for sequence-independent detection of a wide range of viral pathogens (including both DNA and RNA viruses), especially emerging virus such as zika virus, in clinical samples. Its sensitivity is comparable with pipelines developed by other research groups from the US (190) and Switzerland (227) when evaluation was carried out on the reference biological

reagent. In addition to providing the diagnostic information, the method can also generate viral pathogen genome sequences that can be used for additional investigation about the origin and spread of the pathogen.

Herpes simplex viruses are large DNA viruses and are a cell-associated virus. As such they can be partially deposited in the pellets during centrifugation. This explains why the centrifugation associated procedures (approach #3 and #4) failed to detect both HSV-1 and HSV-2 in the virus mixture.

In summary, here I show that a metagenomic approach incorporating nuclease (DNase and RNase) treatment without centrifugation is a sensitive method for sequence-independent detection of a wide range of viruses in clinical samples. This method will be used for analysis of clinical samples. The results of these analyses are presented in subsequent chapters.

Chapter 3: Viruses in patients presenting with community-acquired sepsis of unknown cause in Thailand and Vietnam

1. Introduction:

A recent etiological study of 1578 patients with CA sepsis, conducted by the Southeast Asia Infectious Diseases Clinical Research Network, reported that the etiology (viruses, bacteria and parasites) was established in only 48% (26). Improving our knowledge about the causative agents of CA sepsis can inform clinical management, whilst active surveillance for novel pathogens in this region is of public health significance. Herein, I use mNGS to characterize the viral contents in clinical samples collected from patients enrolled in the aforementioned etiological study of sepsis of unknown etiology across Southeast Asia between 2013 and 2015 (26).

2. Materials and Methods:

2.1. Clinical specimens and patient data:

Clinical specimens and patient data used for mNGS analysis were derived from an etiological study of CA sepsis conducted at multiple hospitals across Indonesia (n=3, Dr. Cipto Mangunkusumo Hospital, Jakarta; Dr. Sardjito Hospital, Yogyakarta; and Dr. Wahidin Soedirohusodo Hospital, Makassar), Thailand (n=4, Queen Sirikit National Institute of Child Health and Siriraj Hospital, Bangkok; Chiang Rai Prachanukroh Hospital, Chiang Rai; and Sappasithiprasong Hospital, Ubon Ratchathani) and Vietnam (n=6, National Hospital of Paediatrics and National Hospital of Tropical Diseases, Hanoi; Hue Central Hospital, Hue; Children's Hospital 1, Children's Hospital 2 and Hospital for Tropical Diseases, Ho Chi Minh City) between 2013 and 2015 (Figure 3.1) (26).

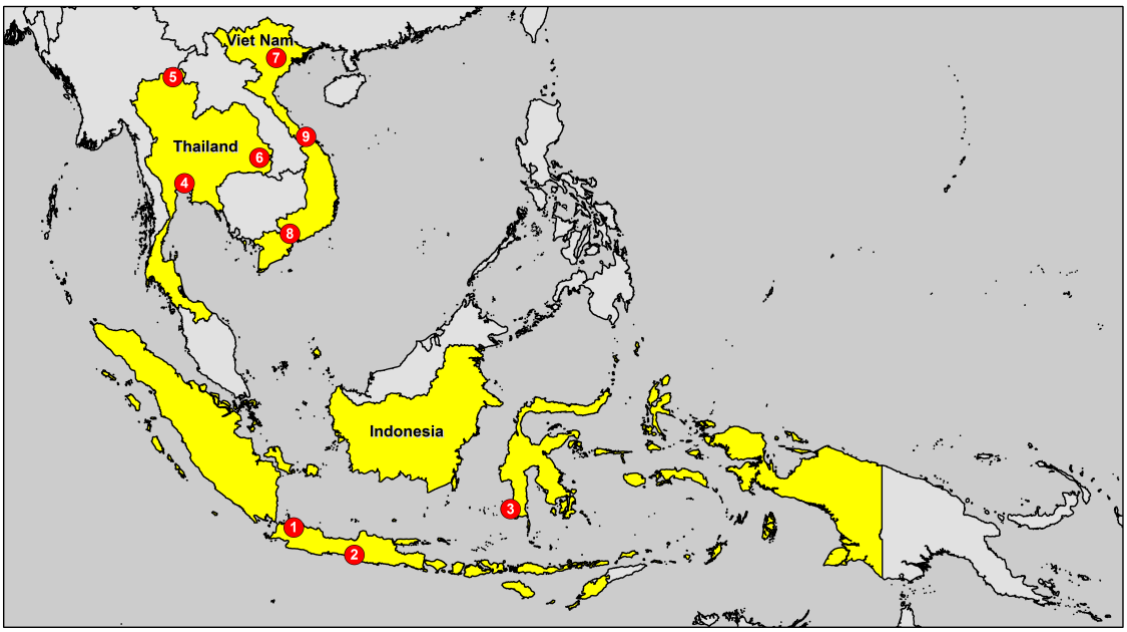


Figure 3.1 Maps of study sites of the original studies. (1) Jakarta, (2) Yogyakarta and (3) Makassar in Indonesia; (4) Bangkok, (5) Chiang Rai and (6) Ubon Ratchathani in Thailand; and (7) Hanoi, (8) Ho Chi Minh City and (9) Hue in Viet Nam. The figure was adapted from "Causes and outcomes of sepsis in southeast Asia: a multinational multicentre cross-sectional study" by Southeast Asia Infectious Disease Clinical Research Network, 2017 (26).

Hospitalized patients with suspected or documented CA infection, fulfilling the diagnostic criteria for sepsis of the 2012 Surviving Sepsis Campaign (adults) (5) or the Pediatric Sepsis Consensus Conference definitions (228) and within 24 h of admission were enrolled (Table 3.1 and 3.2) (26).

Table 3.1 Inclusion criteria used by the original study for patient enrollments

- Age ≥30 days old and weighing at least 3 kg or more on the day of enrollment into the study
- Required hospitalization as decided by the attending physician
- Documented by attending physician that an infection is the primary cause of illness leading to the hospitalization. These can be infections due to any pathogens (bacteria, viruses, fungi and parasites).
- Presence of Systemic Inflammatory Response Syndrome (SIRS)#
- Informed Consent has been obtained

Note: #Systematic response syndrome is described in detail in Table 3.2

Table 3.2 Presence of systematic response syndrome used in the original study

For adults (≥ 18 years old), any combination of a minimum of any 3 of the following 20 parameters	For pediatric patients (>30 days old and <18 years old), all of the 3 following symptoms:
• Fever or hypothermia (Core body temperature defined as >38.3°C or <36.0°C) • Tachycardia (heart rate >90 beats per minute) • Tachypnea (respiratory rate >20 per minute) • Arterial hypotension (systolic blood pressure (SBP) <90 mmHg, mean arterial pressure (MAP) <70mmHg, or SBP decrease >40 mmHg) • White blood cell (WBC) >12,000 u/L or <4000 u/L or immature forms >10% • Platelet count <100,000 u/L • Altered mental status with Glasgow Coma Score (GCS) <15 • Hypoxemia (Pulse Oximetry Level <95) • Ileus • Significant edema or positive fluid balance • Decreased capillary refill or mottling • Hyperglycemia (plasma glucose >140 mg/dL) in the absence of diabetes • Plasma C-reactive protein >2 SD above the normal value • Plasma procalcitonin > 2 SD above the normal value • Arterial hypoxemia (PaO2 / FIO2 <300) • Acute oliguria (urine output <0.5 mL/kg/hr or 45mmol/L for 2 hours) • Creatinine increase >0.5 mg/dL • INR >1.5 or a PTT >60 seconds • Plasma total bilirubin >4 mg/dl or 70 mmol/L	• Fever or hypothermia (rectal temperature defined as >38.5°C or <35.0°C [or equivalent]) • Tachycardia (heart rate >2 SD above the normal value for age). This could be absent in hypothermic subject. • Tachypnea (respiratory rate >2 SD above the normal value for age) AND at least one of the following parameters: • Altered mental status (e.g., drowsiness, poor quality of cry, poor reaction to parent stimuli, and poor response to social overtures) • Systolic blood pressure <2 SD below the normal value for age OR narrow pulse pressure (<20 mmHg) OR poor perfusion (capillary refill >2 sec) • Hypoxemia (Pulse Oximetry Level <95) • White blood cell >15,000 u/L or <5,000 u/L or immature forms >10%.

A total of 1582 patients were enrolled (n=750 each from Vietnam and Thailand, and 82 from Indonesia) (Figure 3.2). Per the study protocol, sera samples were collected from all patients; additional samples including pooled nasal and throat swabs, cerebrospinal fluid and stools were collected when clinically indicated. After collection, all clinical samples were stored at -80^{o}C. Additionally, information about demographics, clinical entities and outcome of the patients was retrieved from a publically available dataset of the original study which was deposited at https://figshare.com/articles/Data_set__Causes_and_outcomes_of_sepsis_in_southeast_Asia_a_multinational_multicentre_cross-sectional_study_NCT02157259_/3486866/1.

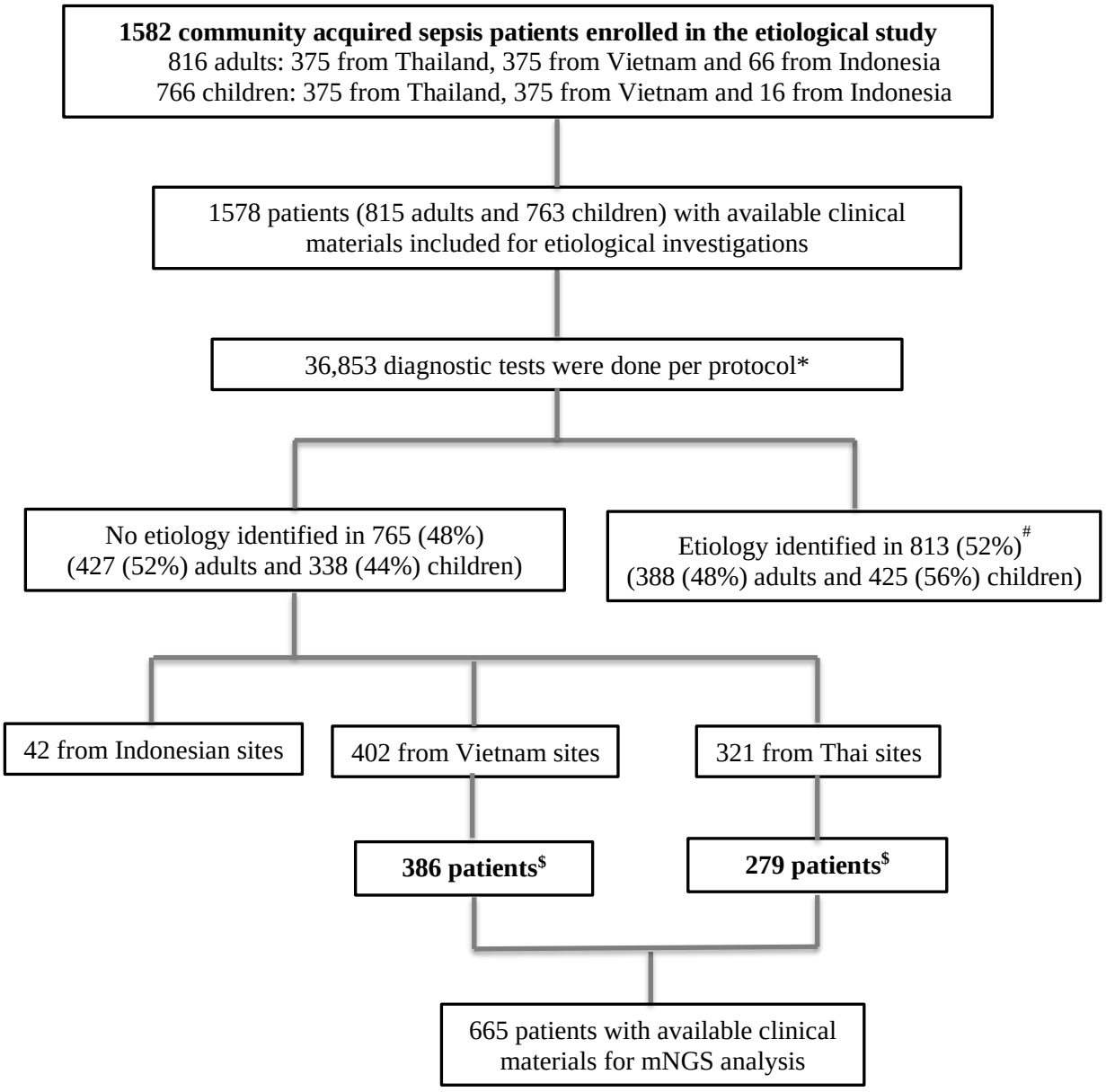


Figure 3.2 Flowchart showing an overview about the diagnostic output of the original study

Notes: *see the original study and Appendix 3 for more details; #the causative agents detected were detailed in the report of the original study; $more details about the analysis of those patients can be found in Figure 3.3.

Clinical samples available for mNGS:

Of 749 Vietnamese patients included in the original study, 402 (54%) had no etiology identified via extensive clinical and reference laboratory work-up (Figure 3.2 and Appendix 3); of whom, 386 (96%) had clinical materials available for additional etiological investigation, and were thus included for viral metagenomic analysis in this study (Figure 3.2) (26). In total, 492 samples (6

CSF, 92 pooled nasal and throat swabs, 384 sera, and 10 stool samples from these 386 patients of unknown etiology were included for analysis.

Of 321 patients from Thailand, there were 279 (87%) patients in whom there was no etiology identified by the original study (Figure 3.2). In total, 351 samples were collected from these 279 patients of unknown origin, including 258 sera, 70 pooled nasal and throat swabs, 22 stools and 1 CSF samples.

Settings

The analysis of samples from Vietnam was conducted at the laboratory of Oxford University Clinical Research Unit in Ho Chi Minh City, Vietnam where I am based. Meanwhile, for the analysis of samples from Thailand, I made two visits with two weeks each to the Mahidol Oxford Tropical Medicine Research Unit in Bangkok, Thailand. Due to the availability of the materials and resources, most samples from Vietnam sites were analyzed individually (n=458) or in pools of multiple samples (n=8) (Figure 3.3). In contrast, because of the resource and time constraints most of samples from Thailand sites were analyzed in pools samples; the only one CSF sample available was analyzed individually (Figure 3.3). Because of these heterogeneities, the results were combined or separately presented when appropriate.

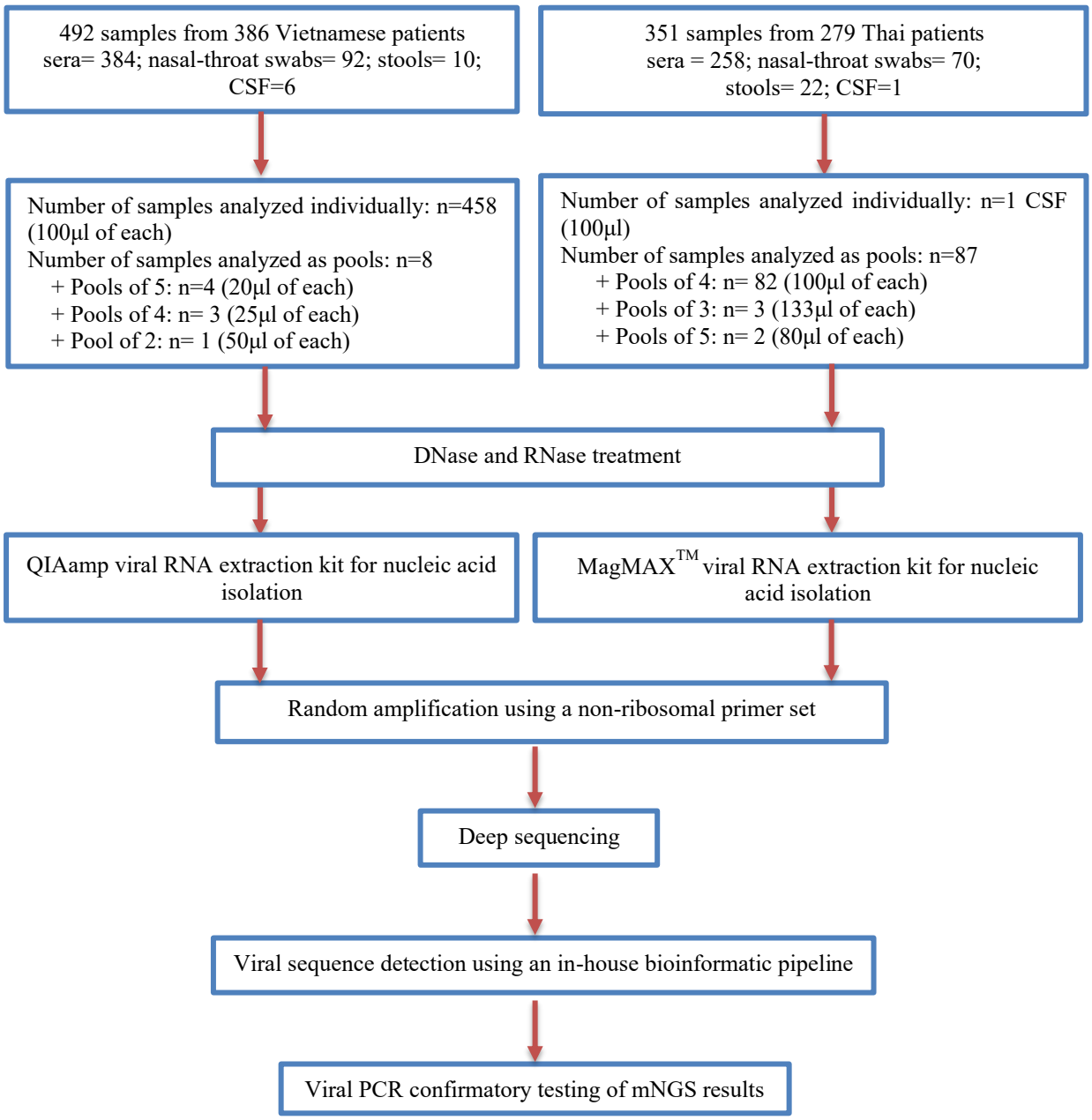


Figure 3.3 Flow chart illustrating how samples were analyzed.

2.2. mNGS assay:

The procedure of mNGS incorporating sample pre-treatment approach #1 (simultaneous DNase and RNase treatment of the original sample without pre-centrifugation) described in Chapter 2 was used to analyse the selected samples, with some modifications for samples from Thailand. Specifically, 400µl of pooled sample was treated with 2U/µl of turbo DNase and 0.4U/µl RNase one (Ambion, Life Technology, Carlsbad, 130 CA, US) at 37 °C for 30 min. The viral NA of treated sample was then isolated by using MagMAX^{TM} viral RNA kit (Ambion, Life Technology, Carlsbad, CA, US), and finally recovered in 50µl of elution buffer. The following dsDNA synthesis, random amplification and sequencing steps are exactly same as described in Chapter 2.

2.3. mNGS data analysis:

Potential viral reads were identified using an in-house viral metagenomic pipeline running on a 36-node Linux cluster as described in detail elsewhere (229). In brief, after duplicate reads and reads belonging to human or bacterial genomes were filtered out, the remaining reads were assembled de novo. The resulting contigs and singlet reads were then aligned against a customized viral proteome database using a BLAST (Basic Local Alignment Search Tool)-based approach. Next, the candidate viral reads were aligned against a non-redundant non-virus protein database to remove any false-positive reads (i.e., reads with expected [E] values higher than those against viral protein databases). Any virus-like sequence with an E value of $\leq 10^{-5}$ was considered a significant hit. Finally, a reference-based mapping approach (Geneious 8.1.5) was employed to assess the levels of identity and genome coverage of the corresponding viruses.

2.4. PCR confirmation of viral reads:

Because of the focus of the present study, specific PCRs were used to confirm the mNGS hits for viral species that are known to be infectious to humans and for recently discovered viruses that have previously been reported in human tissues but remain of uncertain clinical

significance. Depending on the availability of the clinical materials, viral specific PCRs were either carried out on leftover NA after mNGS experiments or on newly extracted NA. A mNGS result was only considered positive if it was subsequently confirmed by a corresponding viral PCR analysis of original NA materials derived from corresponding individual samples. All PCR primers and probes used were either derived from previous publications or newly designed based on the sequences generated by mNGS (Table 3.3).

For Thailand specimens, confirmatory PCR was firstly performed on leftover NA of pooled samples. And if positive, individual samples of positive pools were then analysed. Accordingly, NA was isolated from the corresponding individual samples using QIAamp viral RNA kit (Qiagen GmbH, Hilden, Germany) for subsequent confirmatory PCR analysis.

Table 3.3 List of primers and probes used for subsequent PCR confirmation experiments

Viruses	Oligo sequence (5'-3')			Sources
	Forward	Reverse	Probe	
Measles	ATTACATCAGGATCCGG	GTATTGGTCCGCCTCATC		(230)
HBV	GGACCCCTGCTCGTGTTACA	GAGAGAAGTCCACCMCGAGTCTAGA	FAM-TGTTGACAARAATCCTCACAATACCRCAGA-TAMRA	Newly designed
Rotavirus	ACC ATC TWC ACR TRA CCC TC	GGT CAC ATA ACG CCC CTA TA	FAM-ATG AGC ACA ATA GTT AAA AGC TAA CAC TGT CAA-BHQ1	(231)
Enterovirus	CCCTGAATGCGGCTAAT	ATTGTCACCATAAGCAGCC	CY5-ACCCAAAGTAGTCGGTTCCG -BHQ3	(232)
Dengue	AAGGACTAGAGGTTAGAGGAGACCC	CGTTCTGTGCCTGGAATGATG	FAM- AACAGCATATTGACGCTGGGAGAGACCAGA-BHQ1	(233)
Dengue 2	CCATACACGCCAAACATGAA	GGGATTTCCTCCCATGATTCC	FAM-AGGGTGTGGATTCGAGAAAACCCATGG-BHQ1	(234)
HIV1	GGTGCGAGAGCGTC	ATGCTRTCATCATYTCTTC		(235)
	ATGGGTRAARGTARTAGAAGAAAAGGG	CTGCCTGRTGYCCYCCCACTA		
HCV	AGACTGCTAGCCGAGTAGYGTTGG	TGCTCATGDTGCACGGTCTACGA	FAM-TTGTGGTACTGCCTGATAGGGTGCTT -BHQ1	Newly designed
PIV 1	ATCTCATTATTACCYGGACCAAGTCTACT	CATCCTTGAGTGATTAAGTTTGATGAATA	CYAN500-AGGATGTGTTAGAYTACCTTCATTATCAATTGGTGATG-DB	(236)
PIV2	CTGCAGCTATGAGTAATC	TGATCGAGCATCTGGAAT	LCRED610-AGCCATGCATTCACCAGAAGCCAGC-BBQ	(236)
PIV3	ACTCTATCYACTCTCAGACC	TGGGATCTCTGAGGATAC	LCRED670-AAGGGACCACGCGCTCCTTTCATC-BBQ	(236)
PIV4	GATCCACAGCAAAGATTCAC	GCCTGTAAGGAAAGCAGAGA	HEX-TATCATCATCTGCCAAATCGGCAA-BHQ1	(236)
Coronavirus OC43	GGTGGYTGGGAYGATATGTTACG	KRTTTGGCATAGCACGATCACA	6-FAM-ATGTTGACAAYCCTGTWCTTATGGGTTGGG-MGBNFQ	(236)
Coronavirus NL63	GCTRAGCATGATTTCTTTACTTGG	CARTYTTKTTCATCAAAGTTACGCA	6-FAM-CAGARTCATTTATGGTAATGTTAGTAGACA-MGBNFQ	(236)
PEV	GGGTGGCAGATGGCGTGCCATAA	CCTRCGGGTACCTTCTGGGCATCC		(237)
	YCACACAGCCATCCTCTAGTAAG	GTGGGCCTTACAACTAGTGTTTG		
Rhinovirus	AGSCTGCGTGGCKGCC	ACACGGACACCCAAAGTAGT	CYAN500-TCCTCCGGCCCCTGAATGYGGCTAAYC-DB	(236)
MPV	AGCTTCAGTCAATTCAACAGAAG	CCTGCAGATGTYGGCATGT	LCRED670-TGTTGTGCGGCAGTTTTCAGACAATGC-BBQ	(236)
FA	GACAAGACCAATCCTGTCACYTCTG	AAGCGTCTACGCTGCAGTCC	LCRED610-TTCACGCTCACCGTGCCCAGTGAGC-BBQ	(236)
FB	TCGCTGTTTGGAGACACAAT	TTCTTTCCCACCGAACCA	CYAN500-AGAAGATGGAGAAGGCAAAGCAGAACT-DB	(236)
RSV	ATGAACAGTTTAACATTACCAAGT	GTTTTGCCATAGCATGACAC	LCRED610-TGACTTCAAAAACAGATGTAAGCAGCTCC-BBQ	(236)
			LCRED610-TTATGACATCAAAAACAGACATAAGCAGCTCAG-BBQ	
ADV	CAGGACGCCTCGGRGTAYCTSAG	GGAGCCACVGTGGGRTT	LCRED670-CGGGTCTGGTGCAGTTTGCCCGC-BBQ	(236)
Saffold virus	CTAATCAGAGGAAAGTCAGCAT	GACCACTTGGTTTGGAGAAGCT		(238)
	CAGCATTTTCCGGCCCAGGCTAA	GCTATTGTGAGGTCGCTACAGCTGT		
Salivirus	CCCTGCAACCATTACGCTTA	CACACCAACCTTACCCCACC		(239)
	ATTGAGTGGTGCAYGTGTTG	ACAAGCCGGAAGACGACTAC		Newly designed
Wu-polyomavirus	TGTTACAAATAGCTGCAGGTCAA	GCTGCATAATGGGGAGTACC		(240)
Human herpesvirus 6	TTTGCAGTCATCACGATCGG	AGAGCGACAAATTGGAGGTTTC		(241)
Human herpesvirus 4	GAGGAATTGCCCTTGCTATT	CCTTAGTGGGCCAGGTTGT	FAM -TCGTCTCCCCTTTGGAATGGC-TAMRA	Newly designed
Human herpesvirus 5	CCAAGCGGCCTCTGATAACCA	GGTCATCCACACTAGGAGAGCAGA	FAM-ATGAAGCGCCGCATTGAGGAGATCT-TAMRA	(242)
Gemycircularvirus	GTGGTAATGGTCGTCGGTATTC	CCTCATCATTCGTAGTAAGCAATCTCA		(243)
	AGTCCTGAATGTTTCCACTCG	CAAGCGTTCCCTCGAAAATGAC		Newly designed
Cyclovirus VN	GAGCGCACATTGAAAGAGCTAAA	TCTCCTCCTTCAATGACAGAAACAAC	FAM-CGADAATAAGGMATACTGCTCTAAAGSTGGCG-BHQ1	(244)
Human pegivirus 2	CGCTGATCGTGCAAAGGGATG	GCTCCACGGACGTCACACTGG	CY5-GCACCACTCCGTACAGCCTGAT-BHQ2	(245)

2.5. Phylogenetic analysis:

Sequence alignment and phylogenetic tree reconstructions of the obtained sequences were carried out using ClustalW alignment and Maximum Likelihood methods available within Geneious 8.1.5 (Biomatters) and IQ tree (246), respectively.

2.6. GenBank accession numbers:

Metagenomics data was deposited to NCBI (GenBank) under the accession number PRJNA526981.

2.7. Ethics:

The study was reviewed and approved by the Institutional Review Boards of collaborating hospitals in Vietnam and Thailand, and the Oxford Tropical Research Ethics Committee (OxTREC), University of Oxford, Oxford, United Kingdom. Written informed consent was obtained from either the participant. Or the participant's patents or legal guardian.

2.8. Contributions from others:

Clinical samples and data collection were carried out by participating clinicians and research staff of the respective collaborating hospitals. Although, I led the experiment design and conduct the most of the laboratory work of the laboratory work, I also received support in conducting some of the PCR confirmatory experiments from Ms Le Nguyen Truc Nhu and Ms Nguyen Thi Thu Hong from Emerging Infections group, OUCRU. The in-house viral metagenomic pipeline was conducted at the lab of Prof Eric Delwart at Blood Systems Research Institute, San Francisco, California, United State, with help from Dr Xutao Deng.

3. Results:

3.1. Demographics, clinical features and outcomes for patients with sepsis of unknown origin:

Baseline characteristics and 28-day mortality of all patients (including those of unknown origin from both Vietnam and Thailand) are presented in Table 3.4. Retrospectively, 180 adult patients (including 25% (54/213) and 73.3% (126/172) of the adult patients from Vietnam and Thailand, respectively) had a SOFA score of ≥2, fulfilling the presently used diagnostic criteria for sepsis in adults as defined by sepsis-3 (6). Unlike sepsis in adults, similar harmonized criteria for pediatric sepsis have not been established (247).

There was considerable homogeneity between groups of patients included and not included for mNGS analysis (Table 3.4). Of the patients with unknown cause and included for mNGS, the most frequent clinical entity was acute respiratory infection followed by systemic infection, diarrhea, and central nervous system (CNS) infection (Table 3.4). Thirty-nine patients (37 adults and 2 children) were recorded as deceased by day 28, accounting for 5.9% of the total patients.

Table 3.4 Demographics and clinical data of CA sepsis patients with unknown origin

Baseline characteristics	Vietnamese patients						Thailand patients					
	Patients of unknown etiologies included for mNGS analysis[#]			Patients not included for mNGS analysis			Patients of unknown etiologies included for mNGS analysis			Patients not included for mNGS analysis		
	Total (n=386)	Adults (n=213)	Children (n= 173)	Total (n=363)	Adults (n=162)	Children (n= 201)	Total (n=279)	Adults (n=172)	Children (n=107)	Total (n=471)	Adults (n=203)	Children (n=268)
Gender (male)	224 (58)	122 (57.3)	102 (59)	204 (56)	84 (41)	120 (59)	156 (56)	100 (58)	56 (52.3)	265 (56.3)	117 (57.6)	148 (55.2)
Age												
<12 months	NA	NA	45 (26)	NA	NA	75 (37.3)	NA	NA	8 (7.5)	NA	NA	38 (14.2)
>=1 – <5 years	NA	NA	100 (57.8)	NA	NA	106 (52.7)	NA	NA	41 (38.3)	NA	NA	133 (49.6)
>=5 – <18 years	NA	NA	28 (16.2)	NA	NA	20 (10)	NA	NA	58 (54.2)	NA	NA	97 (36.2)
>=18 – <40 years	NA	94 (44.1)	NA	NA	68 (42)	NA		30 (17.4)	NA		50 (24.6)	NA
>=40 – <60 years	NA	67 (31.5)	NA	NA	60 (37)	NA		55 (32)	NA		64 (31.5)	NA
>=60 years	NA	52 (24.4)	NA	NA	34 (21)	NA		87 (50.6)	NA		89 (43.8)	NA
Geographic location												
North Vietnam	123(32)	68 (32)	55 (32)	127 (35)	57 (35)	70 (34)						
Central Vietnam	141(37)	79 (37)	62 (36)	108 (30)	46 (28)	62 (31)						
South Vietnam	122(32)	66 (31)	56 (32)	128 (35)	59 (37)	69 (34)						
Chang Rai							83 (29.7)	68 (39.5)	29 (27.1)	167 (35.5)	80 (39.4)	87 (32.5)
Bangkok							97 (34.8)	45 (26.2)	38 (35.5)	153 (32.5)	57 (28.1)	96 (35.8)
Ubon Ratchathani							99 (35.5)	59 (34.3)	40 (37.4)	151 (32)	66 (32.5)	85 (31/7)
SOFA score[$]												
<=1	NA	159 (75)	NA	NA	87 (53.7)	NA	NA	46 (26.7)	NA	NA	51 (25.1)	NA
>=2	NA	54 (25)	NA	NA	75 (46.3)	NA	NA	126 (73.3)	NA	NA	152 (74.9)	NA
Clinical presentation*												
Respiratory infection	158(41)	97 (45)	61 (36)	212 (58)	70 (43)	142 (71)	109 (39.1)	72 (41.9)	37 (34.6)	204 (43.3)	81 (39.9)	123 (45.9)
Diarrhea	36 (9)	25 (12)	11 (6)	15 (4)	10 (6)	5 (2)	36 (12.9)	20 (11.6)	16 (15)	62 (13.2)	30 (14.8)	32 (11.9)
CNS infection	40 (10.5)	8 (4)	32 (18)	42 (12)	14 (9)	28 (14)	46 (16.5)	24 (14)	22 (20.6)	66 (14)	32 (15.8)	34 (12.7)
Systemic infection	152 (39.5)	83 (39)	69 (40)	94 (26)	68 (42)	26 (13)	59 (21.1)	35 (20.3)	24 (22.4)	72 (15.3)	39 (19.2)	33 (12.3)
Respiratory and diarrhea							29 (10.4)	21 (12.2)	8 (7.5)	67 (14.2)	21 (10.3)	46 (17.2)
28-day mortality												
Yes	10 (2.6)	8 (3.7)	2 (1)	16 (4)	9 (5)	7 (3)	29 (10.4)	29 (16.9)	0	32 (6.8)	32 (15.8)	0
No	373 (96.6)	203 (95.3)	170 (98)	337 (93)	149 (92)	188 (94)	236 (84.6)	140 (81.4)	96 (89.7)	424 (90)	170 (83.7)	254 (94.8)
Unknown	3 (<1)	2 (1)	1 (<1)	10 (3)	4 (3)	6 (3)	14 (5)	3 (1.7)	11 (10.3)	15 (3.2)	1 (0.5)	14 (5.2)

Notes to Table 3.3: [#]Data were presented as n (%); NA: not applicable; [$]only available for adult patients; *defined based on major clinical symptom: Acute respiratory infection was defined as manifestation of at least one respiratory symptom for no longer than 14 days. Acute diarrhea was defined as diarrhea for no longer than 14 days. Acute CNS infection was defined as manifestation of CNS symptoms for no longer than 14 days or presence of signs of CNS infection on admission. Systemic infection was defined as absence of acute respiratory infection, acute diarrhea and acute CNS infection

3.2. An overview of viral-like sequences detected by mNGS:

Six MiSeq runs (5 for samples from Vietnam and 1 for samples from Thailand) were conducted. Subsequently, a total of over 26 million reads; median (reads/sample): 432,682; range: 540 – 1,916,73 (Appendix 4A) were obtained from 8 pools and 458 individual samples from Vietnam. As for Thai patients, over 24 million reads (median number of reads per samples (range): 139,142 (6,508 – 999,198)) was generated from 87 pools and one single sample (Appendix 4B). Despite the inclusion of a nuclease digestion step prior to NA isolation, viral reads only accounted for a small proportion of total reads, ranging from 168,028 (2.5%) to 287,307 (8.4%) reads/run for Vietnamese samples, and 109,472 (0.44%) reads/run for Thai samples.

Evidence of sequences related to 47 viral species belonging to 21 families were detected in the samples included for analysis (details below). The detected viruses included those known to cause human infections, those with unknown pathogenicity, and viruses that have previously been reported to be contaminants found in mNGS datasets or have not been reported in human samples as detailed below. Additionally, co-detection of ≥2 viruses in the same samples/patients was recorded in 16 patients (Table 3.5). None of the 10 Vietnamese fatal cases had a viral etiology identified by mNGS. Two of 29 Thailand fatal cases were detected with EBV sequences.

Table 3.5 Co-detection of ≥2 viruses in the same samples/patients

	Detected in	Serum	Pooled nasal-throat swabs	Stool
Vietnamese patients	1 Adult	HBV		HBV and Measles virus
	1 Child		Enterovirus, Influenza A and Cytomegalovirus	ND
	1 Child	Cyclovirus VN and Gemycircularvirus	Cytomegalovirus	ND
	1 Child		Enterovirus and Human rhinovirus A	ND
	1 Child		Enterovirus and Epstein Barr virus	ND
	1 Child	Enterovirus	Cytomegalovirus and Epstein Barr virus	ND
	1 Adult	Human immunodeficiency virus, Hepatitis C virus and Human Pegivirus 2	ND	ND
	1 Adult	HBV and Dengue	ND	ND
	1 Adult	ND	ND	Measles and Salivirus A
	1 Child	ND	Cytomegalovirus and Human respiratory syncytial virus	ND
	1 Child	ND	Cytomegalovirus and Human mastadenovirus	ND
	1 Child	ND	Human herpesvirus 6 and Saffold virus	ND
	1 Child	ND	Enterovirus and Human metapneumovirus	ND
Thailand patients	1 Adult		Human alphaherpesvirus 1, Epstein Barr virus	
	1 Child		Epstein Barr virus	Human mastadenovirus
	1 Child		Human mastadenovirus	Human mastadenovirus

3.2.1. Detection of viruses known to cause human infection:

Samples from Vietnam:

Of 466 samples including 458 single samples and 8 pools, NA sequences of 21 viral species known to be infectious to humans were detected in 137 (137/466, 29%) clinical samples from 125 (125/386, 32%) individuals by viral metagenomics. The detection rate was reduced to 12.8% (60/466) samples from 13.4% (52/386) of 386 patients included for mNGS after specific PCR confirmatory testing. There was a significant difference in the number of viral reads generated by mNGS between the groups of samples that were subsequently PCR positive and

negative (Appendix 5), while the total number of reads obtained was similar between the two groups (median (range): 493,794 (11,076 – 1,203,206) vs. 461,486 (16,470 – 1,770,372), p=0.58). The number of reads per samples in the group of samples in which a virus was found by mNGS and subsequently confirmed by PCR was significantly higher than that in the group without a virus found (median (range): 493,794 (11,076 – 1,203,206) vs. 365,974 (540 – 1916732), p=0.004), suggesting that the diagnostic yield of mNGS is dependent on the sequencing depth (i.e. the number of reads generated per sample).

Of the detected viruses, EV was the most common virus (14/386, 3.6%) followed by HBV (9/386, 2.3%), CMV (9/386, 2.3%), human rhinovirus (HRV) (5/386, 1.3%), EBV (5/386, 1.3%), rotavirus (3/386, 0.7%) (Figure 3.4A). The detailed information about the number of viral reads and genome coverage is summarized in Appendix 6.


A
Human pegivirus 2
WU-polyomavirus
Salivirus A
Saffold virus
Influenza B virus
Influenza A virus
Human coronavirus
Human metapneumovirus
Human immunodeficiency virus
Cyclovirus-VN
Human mastadenovirus
Human parechovirus
Human herpesvirus 6
Human parainfluenza virus
Human rhinovirus
Hepatitis C virus
Epstein Barr virus
Measles virus
Rotavirus A
Gemycircularvirus
Dengue virus
Human cytomegalovirus
Human respiratory syncytial virus
Hepatitis B virus
Enterovirus
Total
1 1
1 1
1 1
1 1
1 1
1 1
1 1
1 1
1 1
1 2
2 2
2 2
1 5
2 5
5 6
2 6
5 7
2 8
3 8
5 9
2 9
9 11
2 19
10 32
14 61
76 201
mNGS results confirmed by PCR
mNGS
0 20 40 60 80 100 120 140 160 180 200 220
Numbers of samples

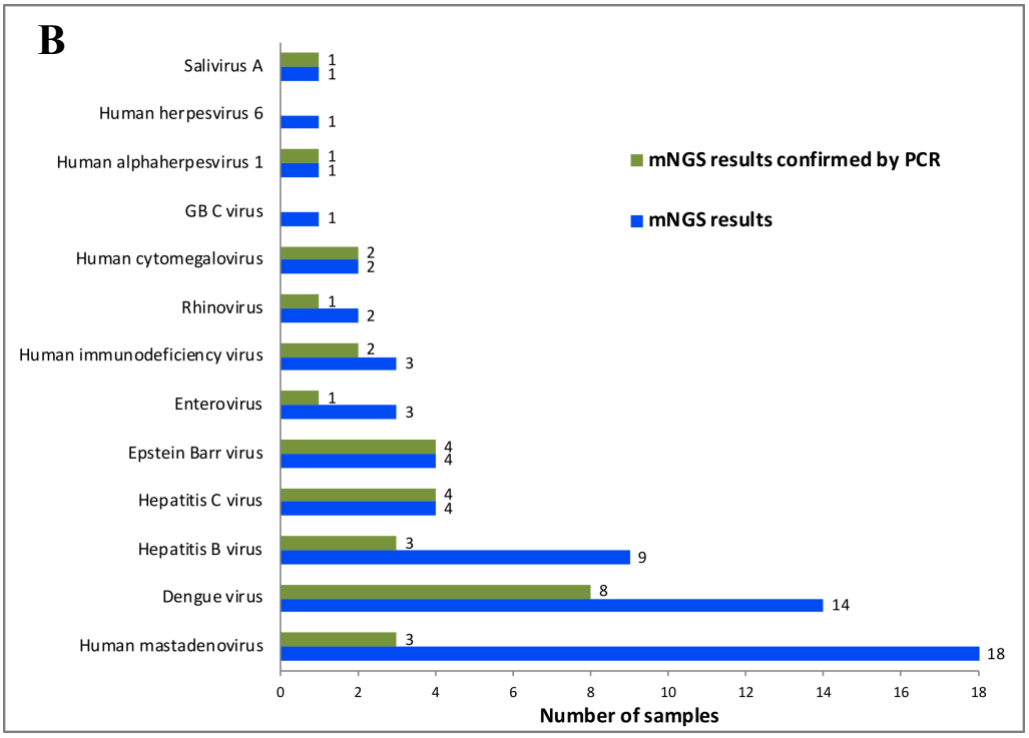


Figure 3.4 Bar chart showing the number of viruses known to be infectious to humans or previously reported in human tissues detected by mNGS followed by PCR confirmation testing. (A) Vietnamese samples, (B) Thailand samples.

Samples from Thailand

Samples were grouped into 87 pools and one single sample for analysis (Figure 3.3.) Evidence of sequences related to 13 human viral species belonging to 6 different families was documented in 63 (71.6%) pools (Figure 3.4B). Only one CSF had no evidence of viral sequences detected by mNGS. After virus-specific PCR confirmatory testing, the detection rate was reduced from 71.6% (63/88) to 34% (30/88) (Figure 3.4B). Subsequent PCR testing of individual specimens of pools with positive PCR results confirmed the presence of 11 virus species in 39/351 (11.1%) of samples from 36/279 (13%) of patients with CA sepsis of unknown cause in Thailand. Among detected viruses, EBV (10/279, 3.6%) and DENV (9/279, 3.2%) were the most common virus detected, followed by HCV (4/279, 1.4%), CMV (4/279, 1.4%), ADV (3/279, 1%), HBV (3/279, 1%) and HIV (2/279, 0.7%) (Figure 3.5).

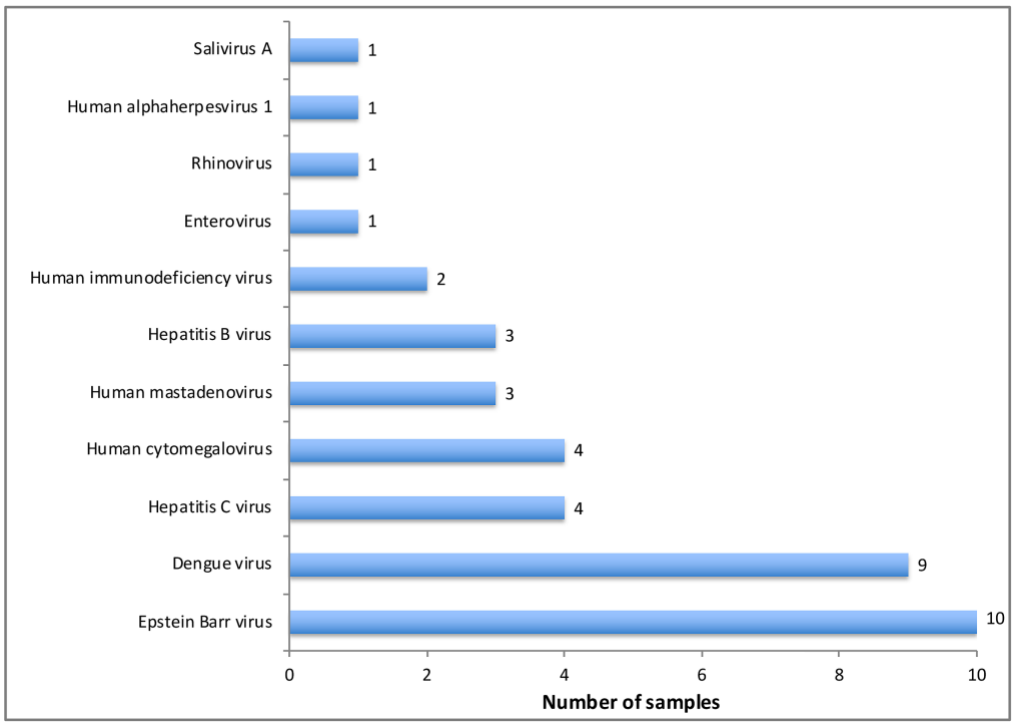


Figure 3.5 Bar chart showing viruses detection by PCR confirmation on individual samples collected from Thailand patients with suspected CA sepsis.

3.2.2. Detection of sequences related to viruses with unknown pathogenicity:

Of 386 Vietnamese patients included for mNGS analysis, sequences related to four recently discovered viruses (gemycircularviruses, WU-polyomavirus, human pegivirus 2 (HPgV-2) and cyclovirus-VN), whose pathogenicity or tropism remains unknown, but genetic materials have previously been reported in human samples, were identified by mNGS in 3.4% of samples. After specific PCR testing, the confirmed number of positive patients was reduced to 2.1% (including gemycircularvirus (5/386, 1.3%), WU-polyomavirus (1/386, 0.26%), HPgV2 (1/386, 0.26%) and cyclovirus-VN (1/386, 0.26%)) (Figure 3.5A). For data generated from Thailand samples, no sequences related to such viruses of unknown pathogenicity was found.

Annellovirus-like sequences were found in the majority of the tested samples, 362/466 (77%) and 62/88 (70%) in Vietnam and Thailand samples, respectively. While sequences related to GB virus C were found in 5 tested samples (4/466 (<1%) individual samples from Vietnam and 1/87 (1%) pooled sample from Thailand), human papillomavirus sequences were found in 1/466

(<1%) Vietnamese sample. Because these viruses are common non-pathogenic infections, they were not subjected to subsequent PCR confirmatory testing.

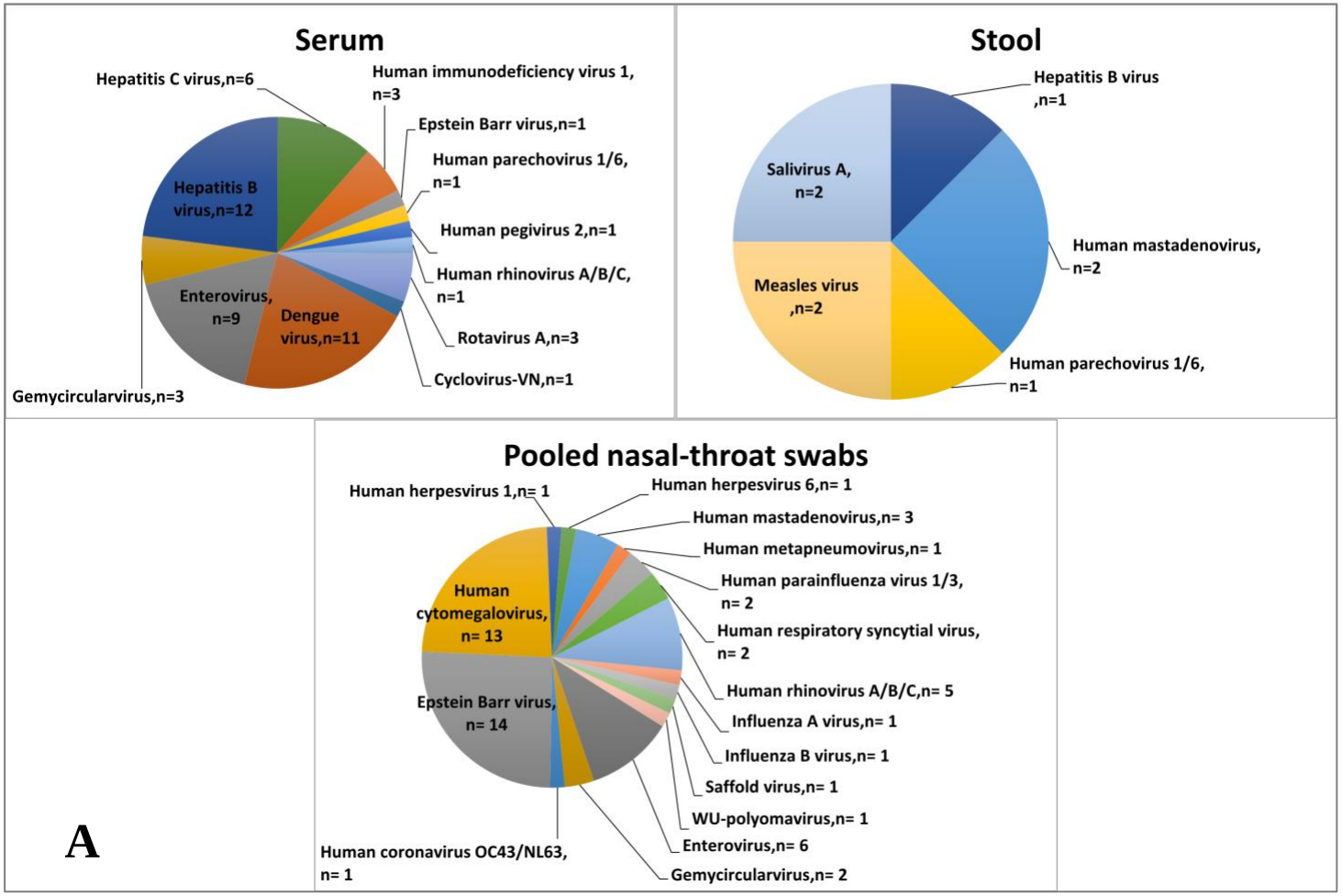
3.2.3. Detection of sequences related to contaminants and or viruses not previously reported in human samples:

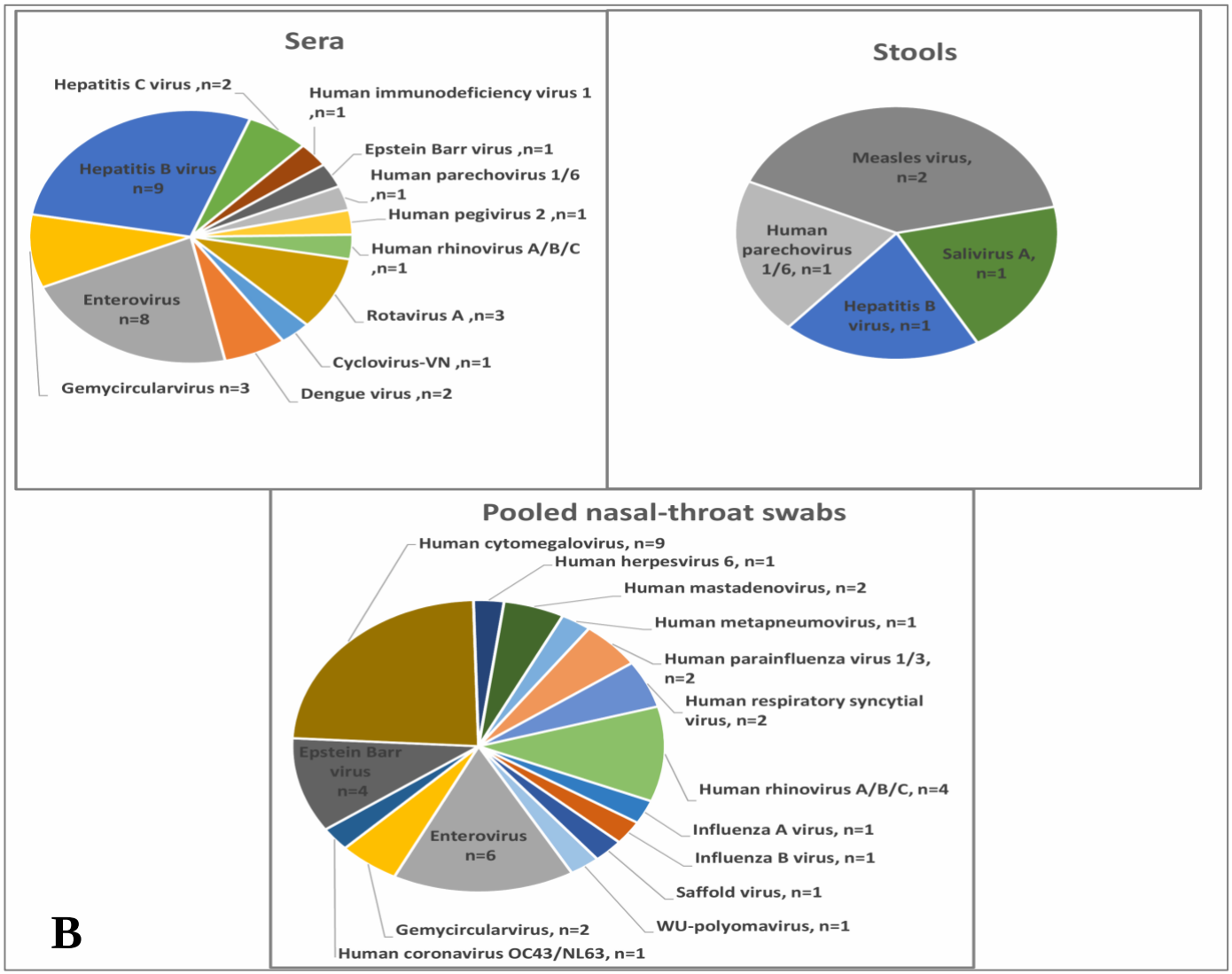
Sequences related to common contaminants of mNGS datasets (including parvo-like hybrid virus (248), and kadipiro virus (249)) were detected in 96 and 5 samples, respectively (Appendix 7). Additionally, sequences related to numerous viruses that have not previously been reported in human tissues were also found (Appendix 7).

3.3. Virus detection by mNGS followed by PCR confirmatory testing in different sample types:

After confirmatory testing, results of individual samples were available for informative pool analysis (i.e. combining data from both sites). The detection rates for human viruses or viruses reported in human tissues were 8% (52/642) for sera/EDTA plasma, 34% (55/162) for pooled nasal-throat swabs and 25% (8/32) for stool samples, while all 7 CSF samples were all negative. More viruses were found in pooled nasal-throat swabs than in other sample types (Figure 3.6A). Of the tested sera, 12 different viral species were detected (including well-established human pathogens; HBV (n=12), DENV (n=11), EV (n=9), HCV (n=6), rotavirus A (n=3), HIV (n=3), human parechovirus (n=1), HRV (n=1) and EBV (n=1) (Figure 3.6A).

Overall, viral richness documented in Vietnamese patient samples was higher than that of Thai patient samples. For instance, influenza A and B viruses were detected in respiratory samples of Vietnamese patients but in none of Thai samples. While HBV and EVs together made up the majority of viruses detected in sera of Vietnamese patients (Figure 3.6B), DENV dominated among viruses detected in sera/EDTA plasma of Thailand patients (Figure 3.6C).


Serum
Hepatitis C virus,n=6
Human immunodeficiency virus 1, n=3
Epstein Barr virus,n=1
Human parechovirus 1/6, n=1
Hepatitis B virus,n=12
Human pegivirus 2,n=1
Human rhinovirus A/B/C, n=1
Enterovirus, n=9
Dengue virus,n=11
Rotavirus A,n=3
Cyclovirus-VN,n=1
Gemycircularvirus,n=3
Stool
Hepatitis B virus ,n=1
Salivirus A, n=2
Human mastadenovirus, n=2
Measles virus ,n=2
Human parechovirus 1/6, n=1
Pooled nasal-throat swabs
Human herpesvirus 1,n= 1
Human herpesvirus 6,n= 1
Human mastadenovirus,n= 3
Human metapneumovirus,n= 1
Human parainfluenza virus 1/3, n= 2
Human cytomegalovirus, n= 13
Human respiratory syncytial virus, n= 2
Human rhinovirus A/B/C,n= 5
Epstein Barr virus, n= 14
Influenza A virus,n= 1
Influenza B virus,n= 1
Saffold virus,n= 1
WU-polyomavirus,n= 1
Enterovirus,n= 6
Human coronavirus OC43/NL63, n= 1
Gemycircularvirus,n= 2
A


Sera
Hepatitis C virus ,n=2
Human immunodeficiency virus 1 ,n=1
Epstein Barr virus ,n=1
Human parechovirus 1/6 ,n=1
Human pegivirus 2 ,n=1
Human rhinovirus A/B/C ,n=1
Rotavirus A ,n=3
Cyclovirus-VN ,n=1
Dengue virus ,n=2
Gemycircularvirus n=3
Enterovirus n=8
Hepatitis B virus n=9
Stools
Measles virus, n=2
Salivirus A, n=1
Hepatitis B virus, n=1
Human parechovirus 1/6, n=1
Pooled nasal-throat swabs
Human cytomegalovirus, n=9
Human herpesvirus 6, n=1
Human mastadenovirus, n=2
Human metapneumovirus, n=1
Human parainfluenza virus 1/3, n=2
Human respiratory syncytial virus, n=2
Human rhinovirus A/B/C, n=4
Influenza A virus, n=1
Influenza B virus, n=1
Saffold virus, n=1
WU-polyomavirus, n=1
Gemycircularvirus, n=2
Human coronavirus OC43/NL63, n=1
Enterovirus n=6
Epstein Barr virus n=4
B

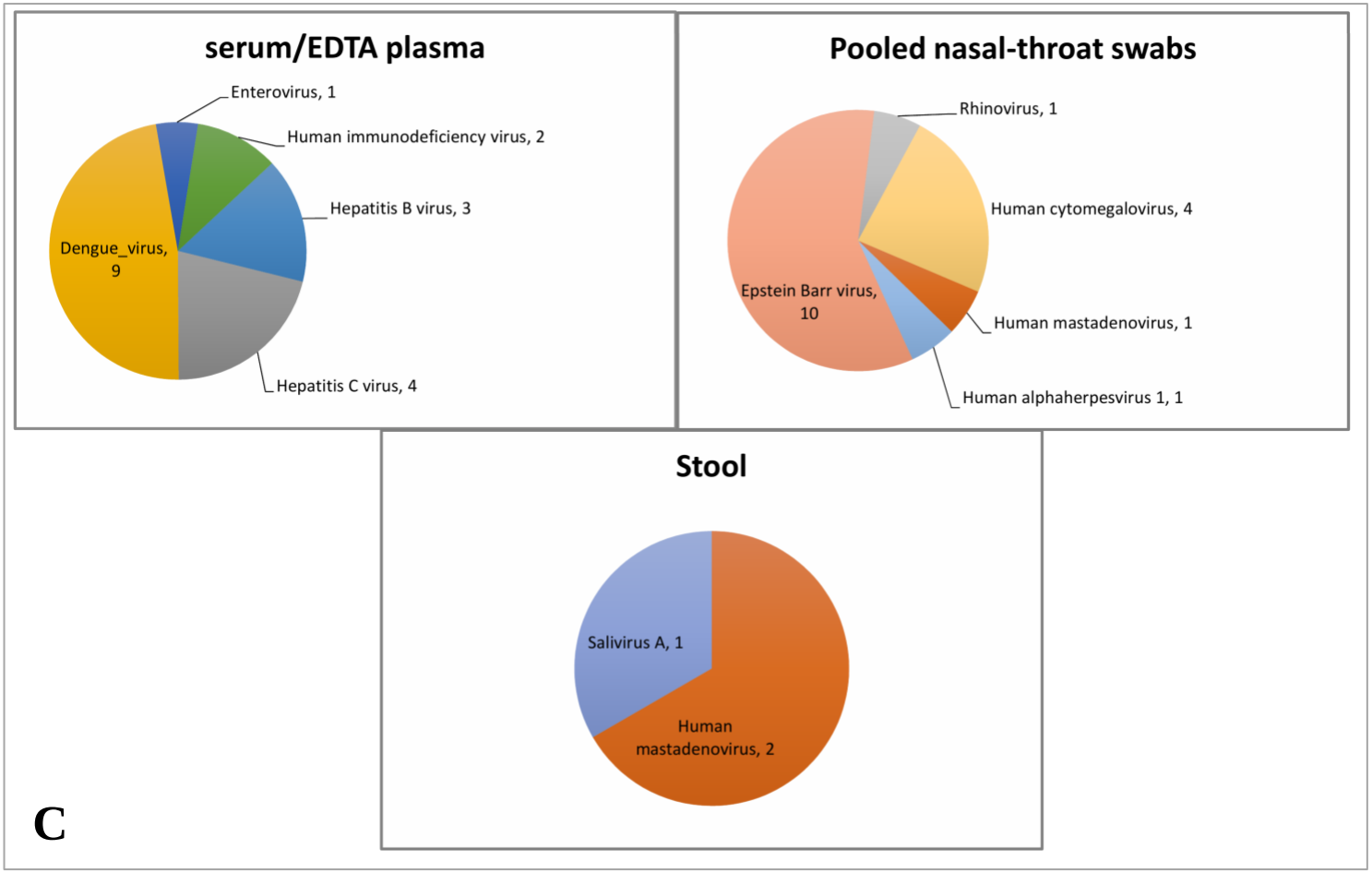


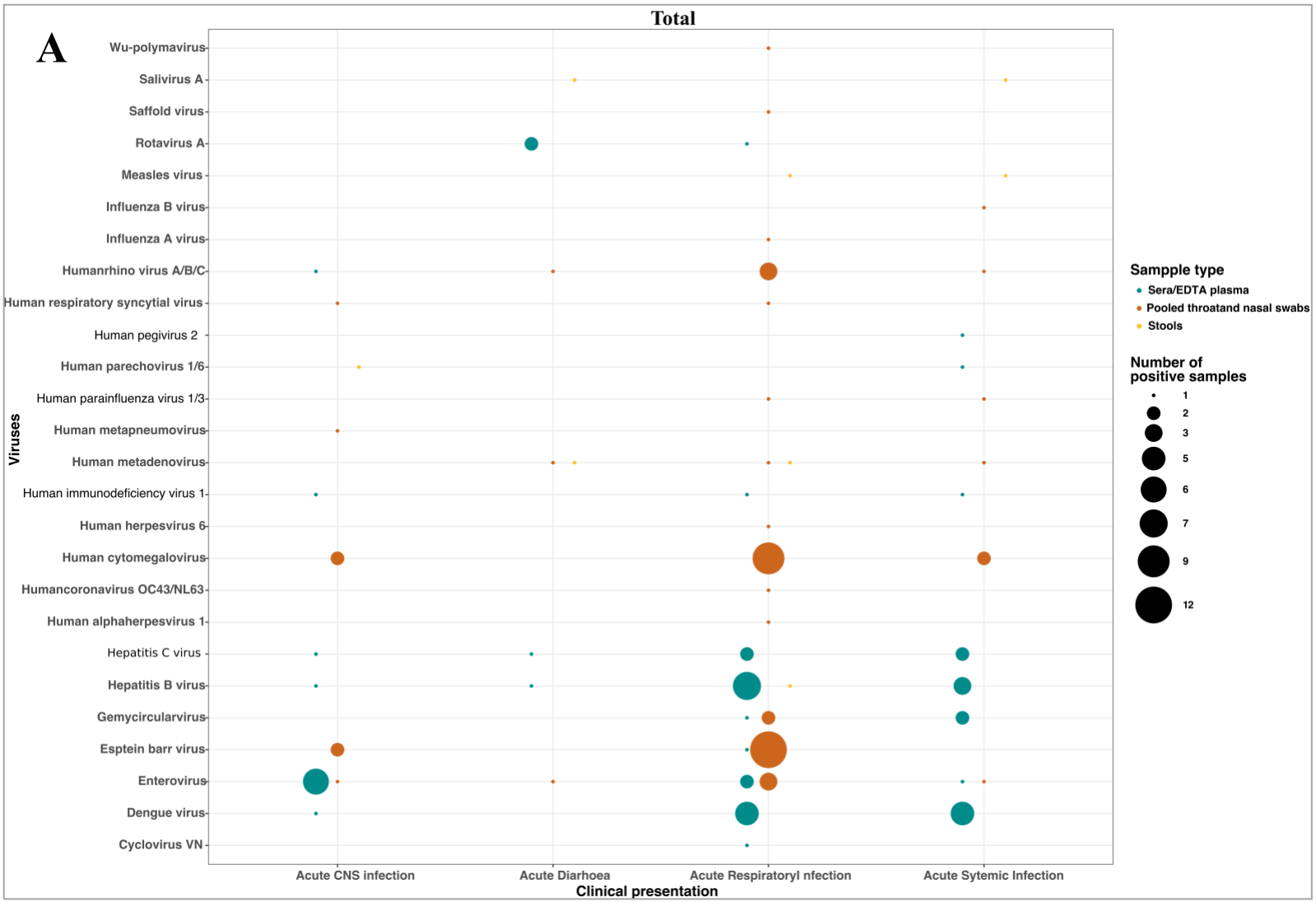
Figure 3.6 The number of viruses detected by mNGS, which were then confirmed by viral specific PCR, in different clinical sample types. (A) Combined data from Vietnam and Thailand samples, (B) Vietnamese samples, (C) Thailand samples.

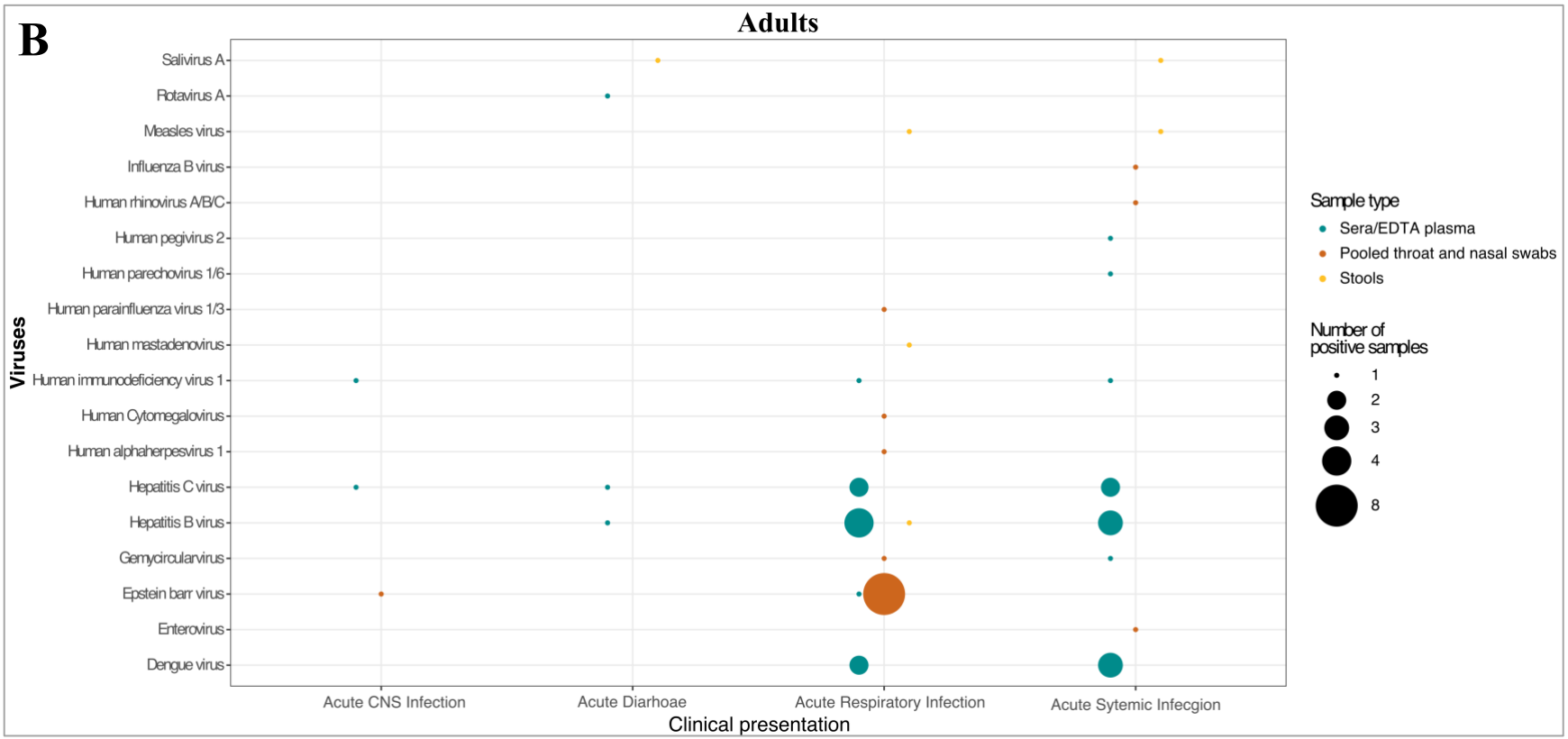
3.4. Viral detection in different patient groups and clinical entities by mNGS followed by PCR confirmatory testing:

Regardless of clinical sample types, pooled analysis for combined data from both sites showed that the highest number of distinct viral infections was recorded in patients presenting with CNS infection (18/86, 21%), followed by patients with respiratory infection (57/306, 18.6%), and patients with diarrhea (10/74, 13.5%) (Figure 3.7A and 3.8A). A similar trend was observed for data from Vietnam (Figure 3.7B and 3.8B). However, of the patients from Thailand, those with virus respiratory infections had more viral species detected and highest diagnostic yield compared to other clinical entities (Figure 3.7C and 3.8C).

Of the Vietnamese patients presenting with CNS infection, picornaviruses were the most common viruses detected (Figure 3.8B), including enterovirus accounting for 7/15 (47%) detected viruses (6 in sera and 1 in a pooled nasal-throat swab), and HRV detected in a serum sample. Of the diarrhea patients, rotavirus, a well-known cause of diarrhea, was detected in blood of three patients (2 children and one adult). Meanwhile, EBV was the predominant virus detected in Thai patients presenting with respiratory illness. Otherwise, the remaining viruses were sporadically detected in samples of patients from both sites.

In terms of age groups, EV, CMV and other respiratory viruses (e.g. respiratory syncytial virus (RSV) and HRV) were more frequently detected in children than in adults (Figure 3.7C). Meanwhile, blood-borne viruses (HIV, HCV and HBV) were more often found in adults than in children (Figure 3.7B). DENV and EBV were commonly found in both age groups. Parechovirus, an established cause of pediatric infections, was detected in one Vietnamese adult presenting with systemic infection.


A
Total
Wu-polymavirus
Salivirus A
Saffold virus
Rotavirus A
Measles virus
Influenza B virus
Influenza A virus
Humanrhino virus A/B/C
Human respiratory syncytial virus
Human pegivirus 2
Human parechovirus 1/6
Human parainfluenza virus 1/3
Human metapneumovirus
Human metadenovirus
Human immunodeficiency virus 1
Human herpesvirus 6
Human cytomegalovirus
Humancoronavirus OC43/NL63
Human alphaherpesvirus 1
Hepatitis C virus
Hepatitis B virus
Gemycircularvirus
Esptein barr virus
Enterovirus
Dengue virus
Cyclovirus VN
Viruses
Acute CNS infection
Acute Diarhoea
Acute Respiratoryl nfection
Acute Sytemic Infection
Clinical presentation
Sampple type
Sera/EDTA plasma
Pooled throatand nasal swabs
Stools
Number of positive samples
1
2
3
5
6
7
9
12


B
Adults
Salivirus A
Rotavirus A
Measles virus
Influenza B virus
Human rhinovirus A/B/C
Human pegivirus 2
Human parechovirus 1/6
Human parainfluenza virus 1/3
Human mastadenovirus
Human immunodeficiency virus 1
Human Cytomegalovirus
Human alphaherpesvirus 1
Hepatitis C virus
Hepatitis B virus
Gemycircularvirus
Epstein barr virus
Enterovirus
Dengue virus
Viruses
Acute CNS Infection
Acute Diarhoae
Acute Respiratory Infection
Acute Sytemic Infecgion
Clinical presentation
Sample type
Sera/EDTA plasma
Pooled throat and nasal swabs
Stools
Number of positive samples
1
2
3
4
8

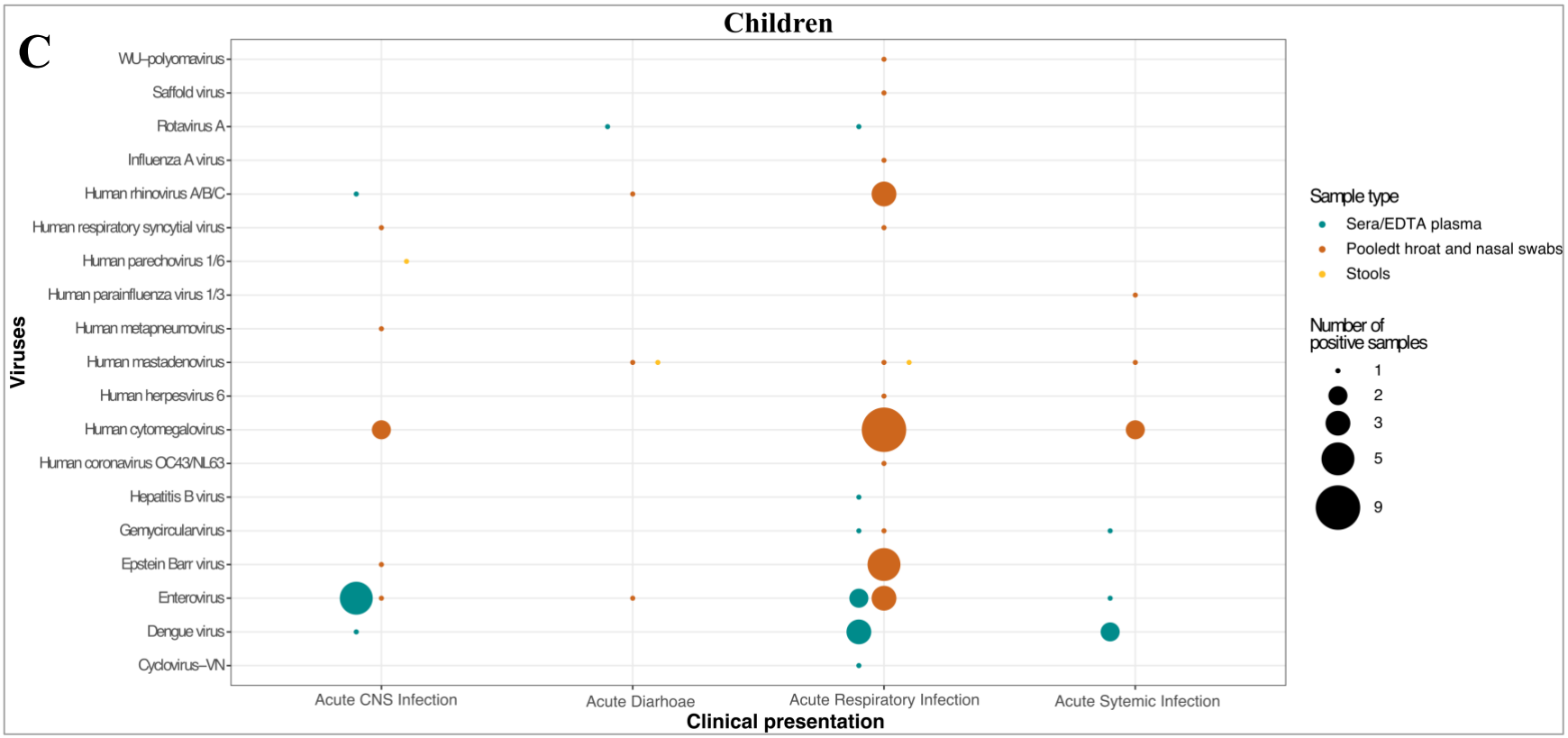
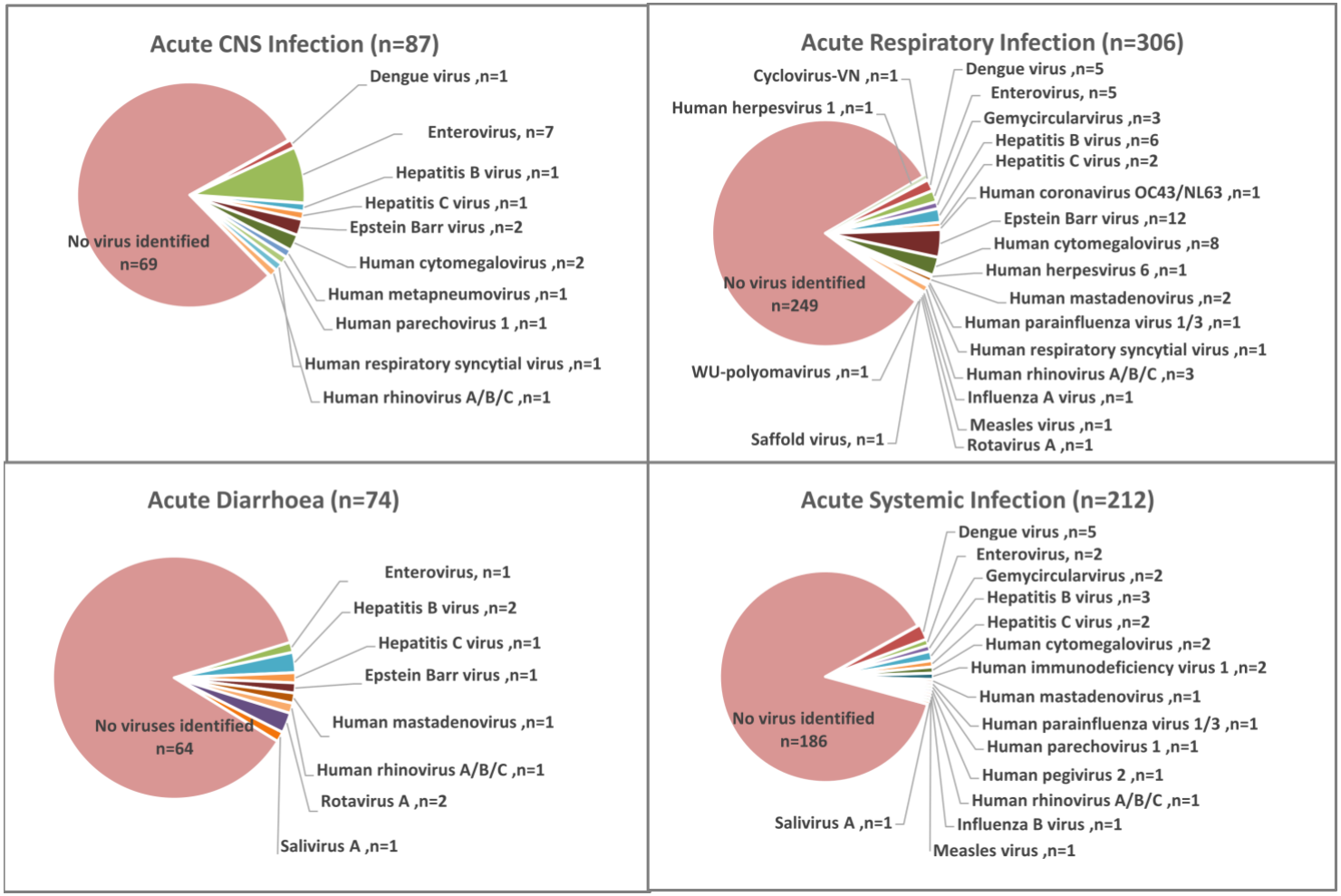
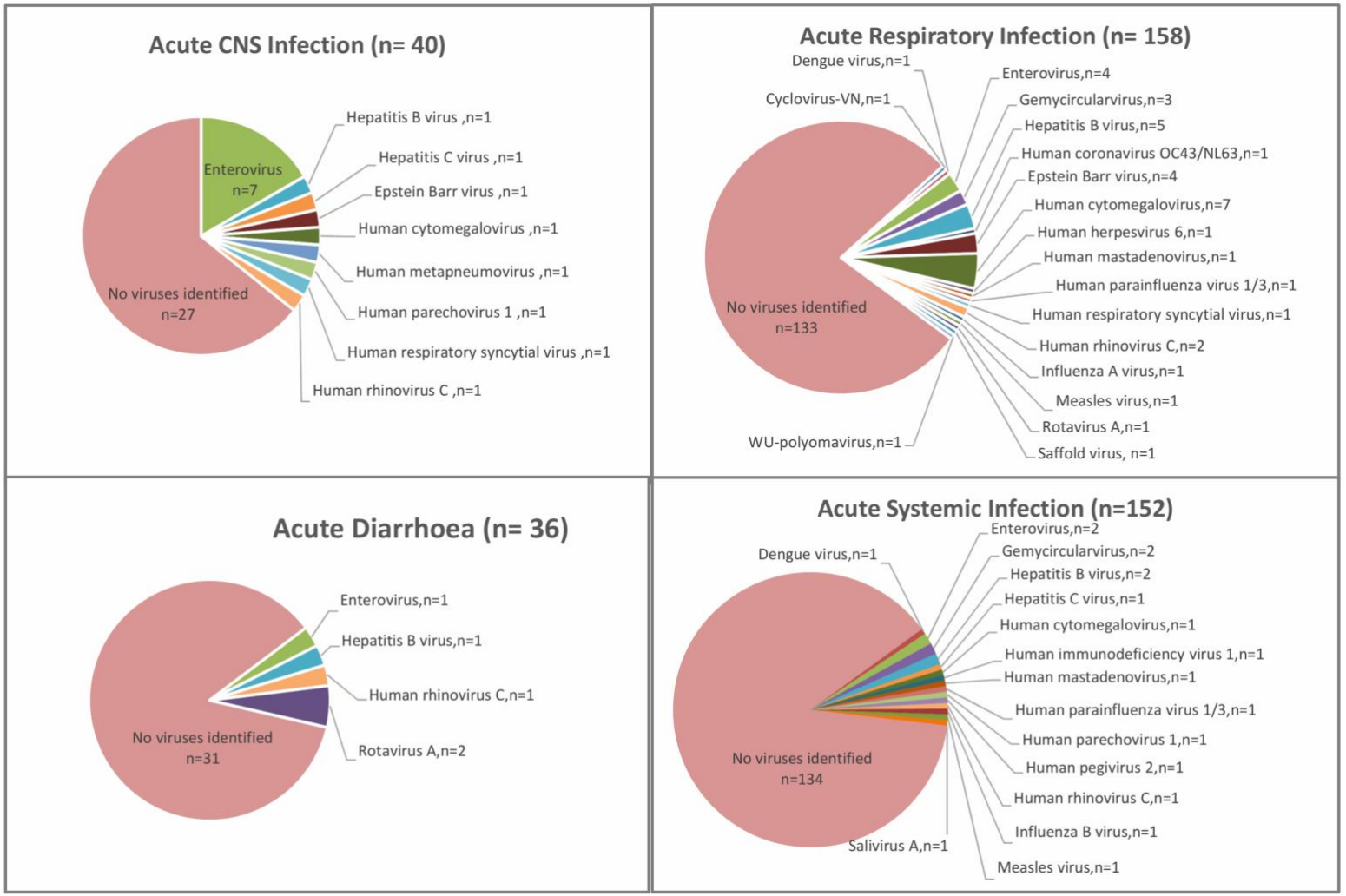


Figure 3.7 The numbers viruses detected by mNGS, which were then confirmed by viral specific PCR, in different patient groups and clinical entities (including in different samples); A) all included patients; B) adults and C) children.

A


Acute CNS Infection (n=87)
Dengue virus ,n=1
Enterovirus, n=7
Hepatitis B virus ,n=1
Hepatitis C virus ,n=1
Epstein Barr virus ,n=2
No virus identified n=69
Human cytomegalovirus ,n=2
Human metapneumovirus ,n=1
Human parechovirus 1 ,n=1
Human respiratory syncytial virus ,n=1
Human rhinovirus A/B/C ,n=1
Acute Respiratory Infection (n=306)
Cyclovirus-VN ,n=1
Human herpesvirus 1 ,n=1
Dengue virus ,n=5
Enterovirus, n=5
Gemycircularvirus ,n=3
Hepatitis B virus ,n=6
Hepatitis C virus ,n=2
Human coronavirus OC43/NL63 ,n=1
Epstein Barr virus ,n=12
Human cytomegalovirus ,n=8
Human herpesvirus 6 ,n=1
Human mastadenovirus ,n=2
Human parainfluenza virus 1/3 ,n=1
Human respiratory syncytial virus ,n=1
Human rhinovirus A/B/C ,n=3
Influenza A virus ,n=1
Measles virus ,n=1
Rotavirus A ,n=1
No virus identified n=249
WU-polyomavirus ,n=1
Saffold virus, n=1
Acute Diarrhoea (n=74)
Enterovirus, n=1
Hepatitis B virus ,n=2
Hepatitis C virus ,n=1
Epstein Barr virus ,n=1
Human mastadenovirus ,n=1
No viruses identified n=64
Human rhinovirus A/B/C ,n=1
Rotavirus A ,n=2
Salivirus A ,n=1
Acute Systemic Infection (n=212)
Dengue virus ,n=5
Enterovirus, n=2
Gemycircularvirus ,n=2
Hepatitis B virus ,n=3
Hepatitis C virus ,n=2
Human cytomegalovirus ,n=2
Human immunodeficiency virus 1 ,n=2
Human mastadenovirus ,n=1
Human parainfluenza virus 1/3 ,n=1
Human parechovirus 1 ,n=1
Human pegivirus 2 ,n=1
Human rhinovirus A/B/C ,n=1
Influenza B virus ,n=1
Measles virus ,n=1
No virus identified n=186
Salivirus A ,n=1

B


Acute CNS Infection (n= 40)
Enterovirus n=7
No viruses identified n=27
Hepatitis B virus ,n=1
Hepatitis C virus ,n=1
Epstein Barr virus ,n=1
Human cytomegalovirus ,n=1
Human metapneumovirus ,n=1
Human parechovirus 1 ,n=1
Human respiratory syncytial virus ,n=1
Human rhinovirus C ,n=1
Acute Respiratory Infection (n= 158)
No viruses identified n=133
Dengue virus,n=1
Cyclovirus-VN,n=1
WU-polyomavirus,n=1
Enterovirus,n=4
Gemycircularvirus,n=3
Hepatitis B virus,n=5
Human coronavirus OC43/NL63,n=1
Epstein Barr virus,n=4
Human cytomegalovirus,n=7
Human herpesvirus 6,n=1
Human mastadenovirus,n=1
Human parainfluenza virus 1/3,n=1
Human respiratory syncytial virus,n=1
Human rhinovirus C,n=2
Influenza A virus,n=1
Measles virus,n=1
Rotavirus A,n=1
Saffold virus, n=1
Acute Diarrhoea (n= 36)
No viruses identified n=31
Enterovirus,n=1
Hepatitis B virus,n=1
Human rhinovirus C,n=1
Rotavirus A,n=2
Acute Systemic Infection (n=152)
No viruses identified n=134
Dengue virus,n=1
Salivirus A,n=1
Enterovirus,n=2
Gemycircularvirus,n=2
Hepatitis B virus,n=2
Hepatitis C virus,n=1
Human cytomegalovirus,n=1
Human immunodeficiency virus 1,n=1
Human mastadenovirus,n=1
Human parainfluenza virus 1/3,n=1
Human parechovirus 1,n=1
Human pegivirus 2,n=1
Human rhinovirus C,n=1
Influenza B virus,n=1
Measles virus,n=1

C

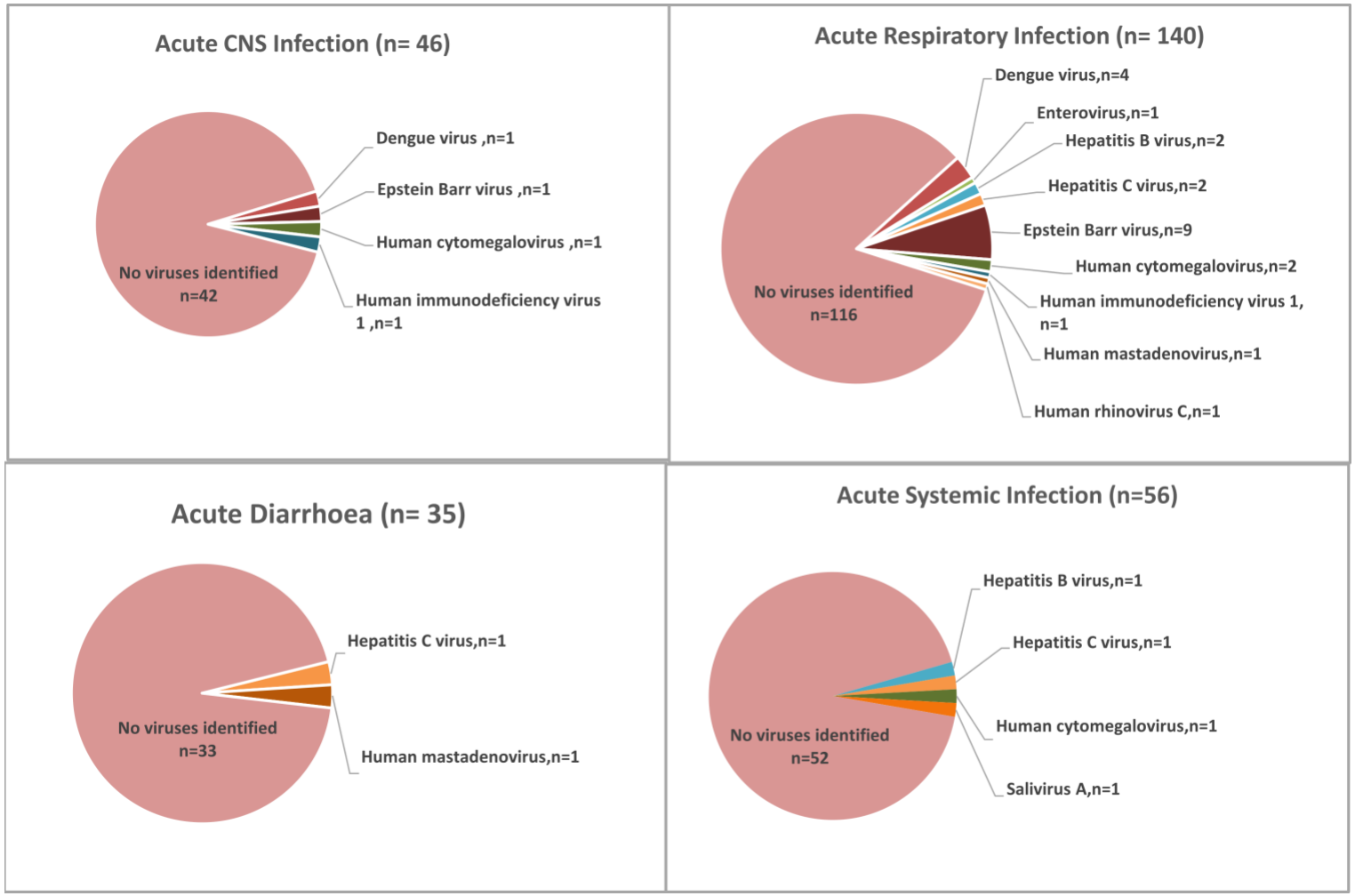


Figure 3.8 Viral detection by mNGS, which were then confirmed by viral specific PCR, in different clinical entities. (A) Combined data from Vietnam and Thailand samples, (B) Vietnamese samples, (C) Thailand samples.

Of the 180 adults with a SOFA score of ≥2, 6 Vietnamese and 13 Thailand patients had a potential viral etiology identified. The detected viruses included EBV (n=6), HBV (n=4), DENV (n=4), measles (n=3), HCV (n=3), rotavirus A (n=2), Gemycircularvirus, Salivirus A and HSV (1 each)) (Table 3.6).

Table 3.6 Viral species found in adult patients with a SOFA score of ≥2

	Patients	Sera/EDTA plasma	Pooled nasal and throat swabs	Stool
Vietnamese patients	1	Rotavirus A	ND	ND
	2	Hepatitis B virus	ND	Measles and hepatitis B virus
	3	Dengue	ND	ND
	4	Gemycircularvirus	ND	ND
	5	ND	Epstein-Barr virus	ND
	6	ND	ND	Measles and Salivirus A
Thailand patients	7	Hepatitis C virus	ND	ND
	8	Hepatitis C virus	ND	ND
	9	Hepatitis C virus	ND	ND
	10	Hepatitis B virus	ND	ND
	11	Hepatitis B virus	ND	ND
	12	Dengue virus	ND	ND
	13	Dengue virus	ND	ND
	14	Dengue virus	ND	ND
	15	ND	Epstein-Barr virus	ND
	16	ND	Human alphaherpesvirus 1, Epstein-Barr virus	ND
	17	ND	Epstein-Barr virus	ND
	18	ND	Epstein-Barr virus	ND
	19	ND	Epstein-Barr virus	ND

Note to Table 3.6: ND: not detected

3.5. Genetic characteristics of detected human viruses in samples of CA sepsis patients from Vietnam and Thailand:

Excluding anellovirus related sequences; mNGS generated sufficient sequence data for informative genetic characterization and phylogenetic inference of EVs, HBV, Dengue virus in 19 samples of CA sepsis patients from Vietnam (n=15) and Thailand (n=4).

All seven complete viral capsid protein 1 (VP1) sequences of EVs were isolated from Vietnamese samples. Phylogenetically, all were classified into six different serotypes of enterovirus A or B (echovirus 3, echovirus 6, echovirus 9, echovirus 16, coxsackievirus A2 and

coxsackievirus A6) (Figure 3.9), supporting previous reports about circulating EVs in Vietnam (250,251).

Seven complete HBV genomes isolated from Vietnam (n=6) and Thailand (n=1) samples. While Vietnamese strains belonged to genotypes HBV-B (n=5) and HBV-C (n=1), Thailand strain belonged to genotype HBV-C (Figure 3.10). Within genogroup HBV-C lineages, Vietnamese and Thailand strains located in two distinct clusters. However, Vietnamese strain showed closely related with previously reported strains from Thailand.

For DENV, the reconstructed tree derived from six complete E gene sequences obtained by mNGS and global representatives (Figure 3.11) showed that the Thailand sequences belonged to serotypes 3 and 4, while Vietnamese sequences were classified into serotype 2. As for other viruses, due to the small number of genomic sequences recovered (Vietnam: gemycircularvirus (2), RSV, influenza B virus, HCV, measles, WU-polyomavirus and cyclovirus-VN (1 each); Thailand: ADV (2), HIV (1)), similar phylogenetic inference was deemed uninformative.

Figure 3.9 Maximum Likelihood tree based on complete VP1 sequences of different enterovirus serotypes illustrating the relatedness between enterovirus serotypes recovered in the present study by mNGS (blue triangles) and representative EV serotypes.

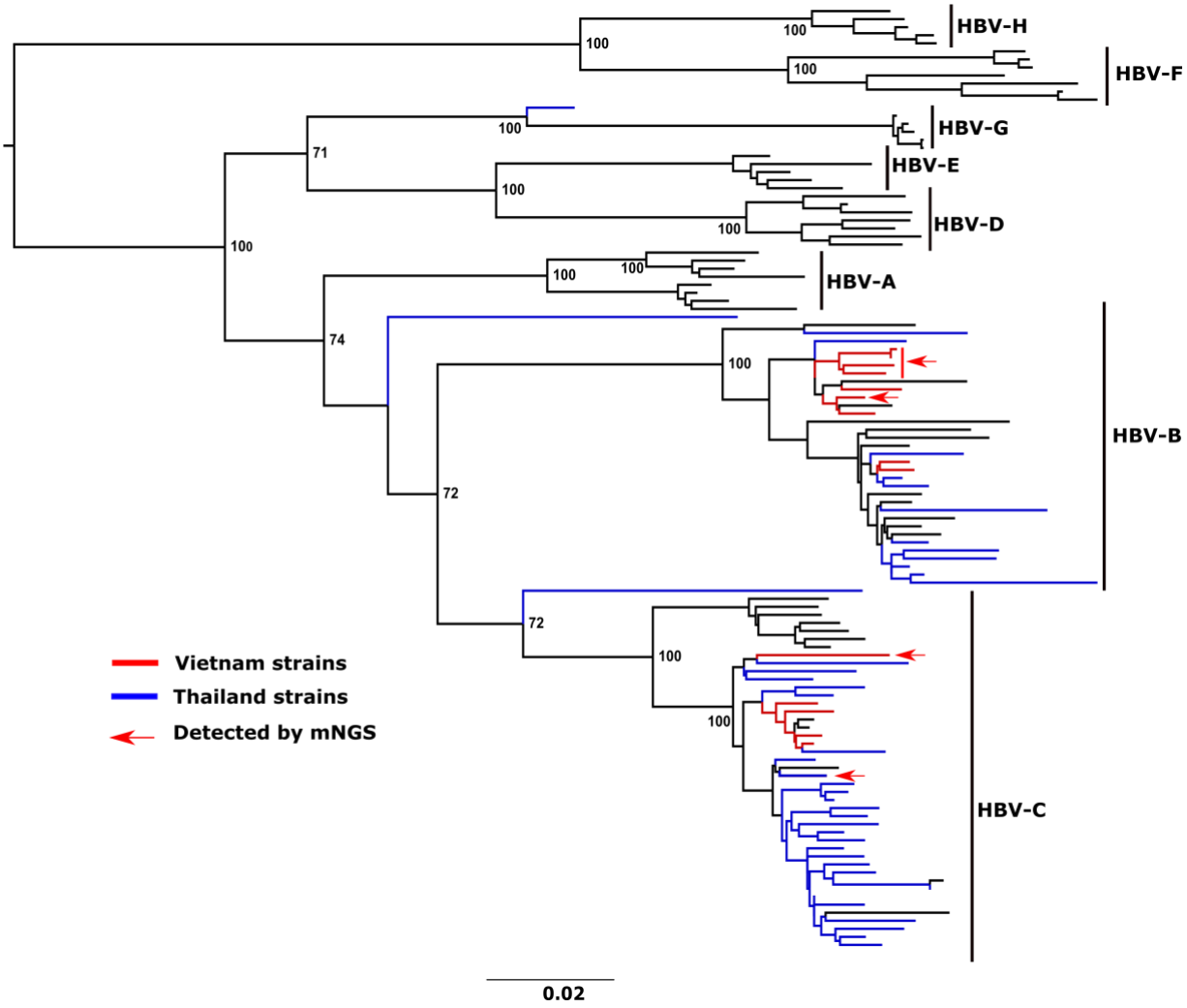


Figure 3.10 Complete coding sequence based Maximum Likelihood tree showing the relationship between Vietnamese (red) and Thailand (blue) strains recovered by mNGS assay (red arrow) with representative HBV genotypes.

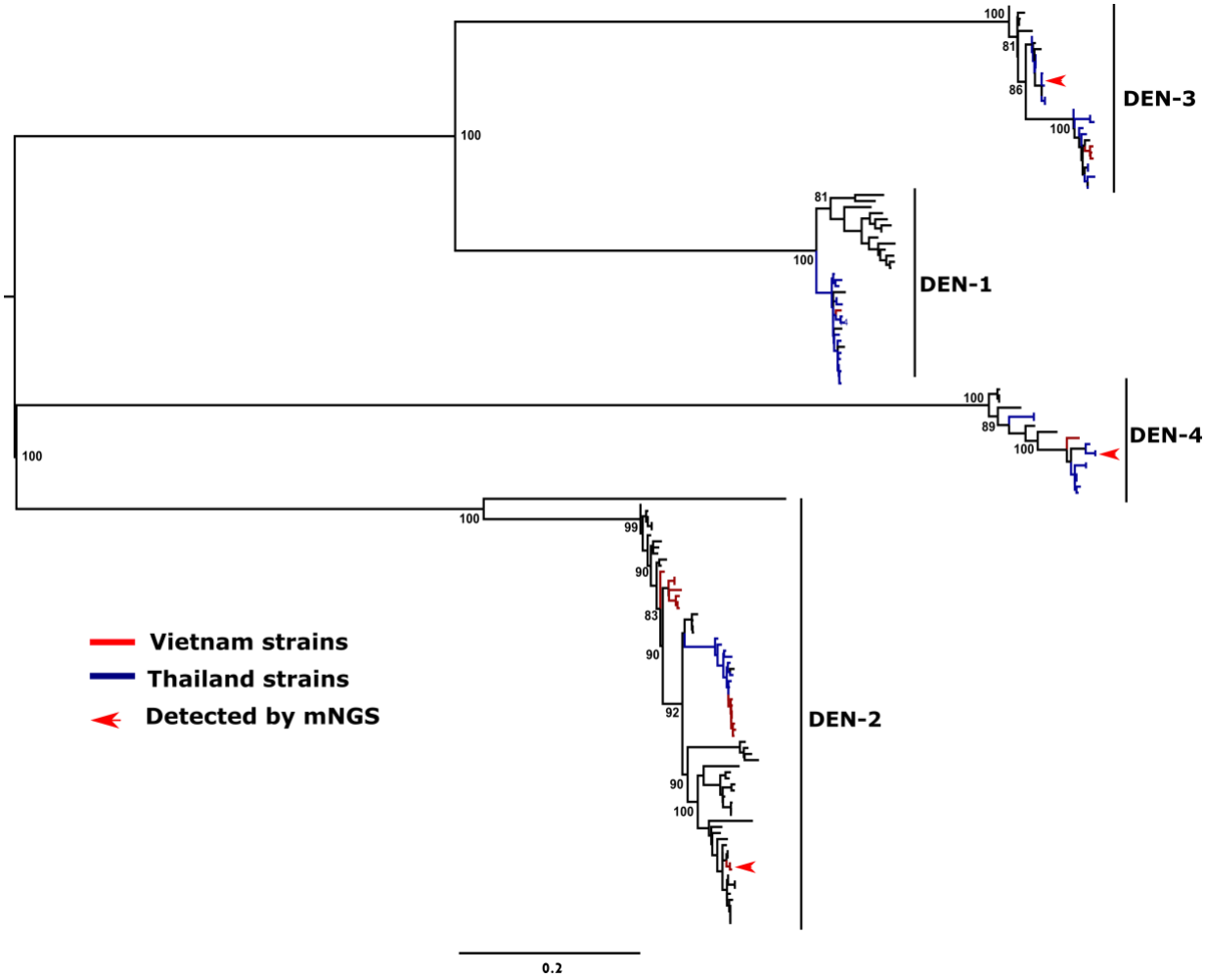


Figure 3.11 Maximum likelihood tree based on complete E gene sequences of dengue virus serotypes illustrating the relatedness between Vietnamese (red) and Thailand (blue) strains recovered by mNGS assay (red arrow) with representative Dengue virus serotypes.

4. Discussion:

I present the results of mNGS to explore the human virome in 665 patients presenting with CA sepsis of unknown cause enrolled in a multicenter observational study across Vietnam (n=386) and Thailand (n=279) from 2013 to 2015. I identified 22 viral species known to be infectious to humans in 90 (13.5%) of 665 patients presenting with CA sepsis of unknown cause. Similar diagnostic yields were also obtained for samples from each site. The study, however, cannot directly impute sepsis causation involving the viruses identified. More specifically, on several occasions, viral detection in non-sterile materials such as respiratory (including EBV and CMV) and stool samples may simply reflect the carriage of such viruses in those bodily compartments rather than a clinical association. Similarly, viral detection (e.g. enterovirus) in blood of cases with asymptomatic infection has previously been reported (252). Additionally, the detection of blood-borne viruses such as HBV, HIV and HCV in sera samples might represent underlying diseases and not the causative pathogens leading to the hospital admission, although the detection of HIV RNA in a serum sample of patients presenting with systemic infection and CNS infection may suggest acute HIV infection. However, together with the clinical and epidemiologic data, the results present a provocative argument for a wide range of viral pathogens that might be associated with CA sepsis in Southeast Asia.

Epidemiologically, my results support previous findings regarding the frequent detection of common viruses in corresponding clinical entities and age groups. For examples, I only found rotavirus in patients with acute diarrhea and RSV, and detected viruses of the *Picornaviridae* family (HRV and EV) mostly in children, while HBV, HCV and HIV mainly are detected in adults. Additionally, I detected parechovirus in blood of an adult presenting with acute systemic infection. Parechoviruses are a well-known cause of disease in children, ranging from acute gastrointestinal/respiratory infections to meningitis, but have increasingly been reported to cause infections in adults (253).

Additionally, my analysis also demonstrated the levels of viral richness and diversity were slightly different between Thailand and Vietnam. For instance, more viral species were found in Vietnamese patients than in Thai patients, despite similar diagnostic yields for both patient groups. Additionally, I found more samples positive for Dengue in Thai patients than in Vietnamese patient, while influenza viruses were only detected Vietnam. Both Thailand and Vietnam are endemic countries for Dengue, whilst influenza viruses are circulating globally. Thus, the differences might simply reflect the association with the fluctuation in prevalence of these viruses in respective countries during the study period.

Non-polio enteroviruses like EV-A71 and EV-D68 have become serious global threats. In fact, EV-A71 has overwhelmed countries of the Asia-Pacific region (including Vietnam) with large outbreaks of severe hand, foot and mouth disease since 1997 (254,255). Recently, EV-D68 has emerged and caused large outbreaks of respiratory infections in the U.S. and is epidemiologically linked with acute flaccid myelitis (109). Collectively, the data presented here combined with results of the original report (26) expand our knowledge about the clinical burden posed by non-polio enteroviruses (HRV and particularly diverse EV serotypes) and parechoviruses in Vietnam.

mNGS detected several recently discovered viruses (saffold virus, salivirus A, Wu-polyomavirus, gemycircularvirus and HPgV-2), representing the first detection of these viruses in Vietnam and Thailand, and adding to the growing literature about the geographic distribution of these newly identified viruses. Saffold virus and salivirus A have been linked to gastrointestinal and/or respiratory infection, respectively (238,256–258), while the former has also been reported to be associated with myocarditis and aseptic meningitis (259,260). Additionally, using a mouse model, studies have shown the neurotropic potential of saffold virus (260–262). The pathogenicity of Wu-polyomavirus, gemycircularvirus and HPgV-2 remains unresolved. Likewise, it is imperative to conduct follow-up studies to determine whether the

detected sequences related to viruses that have not previously been reported in human tissues are derived from other sources, and whether the respective viruses are infectious to humans.

The results of the present investigation also emphasize the utility of sera samples for assessing the etiology of sepsis. Indeed, viruses of the family *Picornaviridae* (enterovirus, rhinovirus, and parechovirus), *Flaviviridae* (DENV) and *Caliciviridae* (rotavirus) were detected by mNGS in the included sera. Notably, as per the design of the original etiological study, sera were not tested for these viruses by PCR (26). Likewise, while it remains unknown why the original study failed to detect common causes of respiratory/enteric infections (influenza A/B virus, EV, etc.) in pooled-nasal swabs by multiplex PCR assays and substantial cases of DENV in serum/plasma samples (26), a slight decrease in sensitivity of the multiplex PCR assays used as compared to that of respective monoplex PCR assays has been reported elsewhere (236).

Virus detection by mNGS is based on the detection of matching viral reads regardless of their number or resulting genome coverage. While few metagenomic studies published to date reported the use of specific PCR to subsequently verify metagenomic results, the failure of viral specific PCR to confirm the original mNGS detections for many patients in the present study may be a consequence of cross-talk (bleed over) contamination occurring as part of the sequencing procedure, a well-documented phenomenon (263–265). An alternative explanation is the low sensitivity, likely attributed to nucleotide mismatches, of some of PCR primers used to confirm infection.

The analytical approaches were slightly different for samples from Vietnam (mostly individual samples were analysed) and Thailand (mostly samples were analysed in pools). Despite these variations, comparable detection rates by mNGS were obtained, 13.4% (52/368) for Vietnamese patients and 13.6% (38/279) for Thai patients. Thus, the data suggest that the absence of human viral pathogens in 87% of 665 patients may be attributed to some possibilities. It might be the case that the sensitivity of current mNGS approach was not sufficient to detect pathogens

presenting in the tested samples with low viral load. This is even more prominent in cases where the number of reads obtained were supposedly insufficient (Supplementary Figure 1), as suggested by the difference in the number of obtained reads between the groups of samples with and without a virus identified. Clearly, future research should address at what level of sequencing depth mNGS based approaches need to achieve in order to reach the required sensitivity, while maintaining the cost effectiveness. Of equal importance is to identify the factors (e.g. sample types and library preparation/sequencing methods) that may affect the sequencing depth (i.e. the number of reads obtained) and the assay sensitivity. Additional possibilities include the presence of sepsis pathogen in non-analyzed tissues, the presence of non-viral pathogens (e.g. bacteria and parasites) in tested specimens, and/or the inclusion of patients with non-infection (e.g. those caused by toxicity with clinical presentations mimicking infections) in the study.

In summary, I report the application of mNGS in patients presenting with CA sepsis of unknown etiology. My results highlight challenges in identifying possible viral culprits in patients with CA sepsis, and that diverse viral agents might be responsible for such devastating conditions in tropical settings like Southeast Asia. Therefore, rigorous testing for a wide range of viral pathogens in samples from different body compartments collected early after symptom onset when viral loads are usually highest is likely to have the greatest yield. Under these circumstances, mNGS is a promising approach because of its capacity to simultaneously detect and genetically characterize viral pathogens in patient samples without the need of prior knowledge about genomic information of the targeted pathogens, thereby enhancing the ability to identify infectious etiologies of sepsis, and facilitating optimal targeted management.

Chapter 4: Detection and Characterization of Human Pegivirus 2 in Vietnam

1. Introduction:

Recently, a new pegivirus species, namely human pegivirus 2 (HPgV-2) or human hepegivirus 1 (herein I use the term HPgV-2), of the genus *Pegivirus* and the family *Flaviviridae* was discovered by two independent research groups in the United States (266,267).

HPgV-2 genome is positive single-stranded RNA with about 9,8kb in length. It consists of a single open reading frame encoding for a multifunctional polyprotein (266). Genetically, the level of nucleotide identity between individual HPgV-2 strains identified to date ranges from 93% to 94.4%, while it shares ≤32% amino acid similarity with the other pegivirus species (266). Existing evidence suggests that HPgV-2 is a blood-borne virus and is more frequently detected in patients with HCV infections, particularly HCV and human HIV co-infection, although detection rates vary between studies and patient groups. In the U.S. study by Berg et al, HPgV-2 was detected in 11/982 (1.1%) of patients with active HCV infections, while HPgV-2 RNA was absent in patients with HIV or HBV infections (n=494 and 488, respectively) as well as in volunteer blood donors, who were HIV, HBV and HCV negative, (n=476) (266). Most recently, HPgV-2 RNA was detected in 0.29% (7/2440) and 3.47% (7/202) of HCV mono-infected patients and HCV/HIV co-infected subjects in China, respectively (268). Among injection drug users, HPgV-2 RNA was detected in 10.9% (17/156) of individuals who were either HCV or anellovirus SEN virus D positive, in the U.S. (269), and in 5.7% (4/70) and 3% (8/270) of HCV and HIV co-infected subjects in Guangdong and Sichuan, respectively, in China (270).

HPgV-2 viremia has been shown to persist for up to 6158 days (267,269) suggesting that HPgV-2 can establish chronic infection. Although factors predisposing to chronic infection are unclear,

codependence between HPgV-2 other viruses, particularly HCV, may play a role analogous to that of HBV and hepatitis delta virus (271).

Given the high burden of HCV and HIV infections worldwide, and the potential clinical significance of HPgV-2, in this chapter I further characterized the HPgV-2 sequences detected in a HIV and HCV co-infected patient by mNGS in Chapter 3. I also aimed to investigate the geographic distribution, genetic diversity, and prevalence of this virus to help prioritize the development and implementation of appropriate intervention strategies.

2. Methods and materials:

2.1. Patients and clinical samples for initial viral metagenomic analysis:

Patient information and clinical samples were derived from a multicenter observational study designed to evaluate the causes of CA sepsis in children and adults in Southeast Asia, as described in Chapter 3 (26). All patients enrolled at sites in Vietnam who did not have an etiology identified via extensive clinical and reference laboratory work-up in the original study were selected for additional viral metagenomic analysis in this study, as described in Chapter 3.

2.2. Clinical samples and viral metagenomic datasets for subsequent human pegivirus 2 screening:

After the initial detection of HPgV-2 in plasma of one patient of the CA sepsis infection cohort, I expanded the testing to other clinical studies and metagenomic datasets to further investigate its prevalence and genetic diversity. These consisted of patients with HCV (n=394), HBV (n=103), hepatitis A virus (HAV) (n=71), HIV (n=78) and healthy control subjects (n=80) (Table 4.1). These patients participated in a clinical trial evaluating the hepatic safety of raltegravir/efavirenz-based therapies in antiretroviral-naive HIV- infected subjects co-infected with HCV.

Table 4.1 Samples and viral metagenomic datasets used for screening of HPgV-2, and screening results

Disease	No.	Screening approach	Detected in	Enrolment period	Setting
Hepatitis C virus and HIV coinfection*	79	HPgV-2 specific PCR and reference-based mapping of obtained viral metagenomics data	5	2010 – 2013	Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam
HIV mono-infection	78	HPgV-2 specific PCR	0	2010 – 2013	
Hepatitis A virus	71	HPgV-2 specific PCR	0	2012 – 2016	
Hepatitis B virus	103	HPgV-2 specific PCR	0	2012 – 2016	Hospital for Tropical Diseases, Ho Chi Minh City, Dong Thap General Hospital, Dong Thap, Khanh Hoa Provincial Hospital, Nha Trang, Dac Lac Provincial Hospital, Dac Lac, Hue National Hospital, Hue; All in Vietnam
Hepatitis C virus*	394	Reference-based mapping of obtained viral metagenomics data	0	2012 – 2016	
Healthy volunteers	80	HPgV-2 specific PCR	0	2010 – 2013	Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam

Note: *Whole-genome sequences of hepatitis C virus were obtained using a viral metagenomics approach (221). The resulting metagenomics datasets were then subjected to a reference-based mapping approach to search for the presence of HPgV-2 sequences.

2.3. HPgV-2 PCR:

A previously described multiplex real time PCR (RT-PCR) targeting two different conserved regions of the HPgV-2 genome (5’ untranslated region (5’UTR) and non-structural 2-3 coding region (NS2/3)) were employed to screen for HPgV-2 in clinical samples (245). In details, the multiplexed RT-PCR was carried out using the SuperScript III One-Step qRT-PCR System with Platinum Taq DNA Polymerase (Invitrogen, Carlsbad, CA, US) and performed in a LightCycler 480 II machine (Roche Diagnostics GmbH, Mannheim, Germany). The PCR reaction contained 5µl of viral RNA, 12.5µl of 2x reaction mix, appropriate concentrations of primers and probes (245) and 0.5µl of enzyme mix. The thermal cycling condition consisted of one cycle of 60^{o}C for 30 min, 95^{o}C for 1min followed by 40 cycles of 95^{o}C for 30s, 60^{o}C for 1min (including fluorescence acquisition) and 72^{o}C for 30s. Details of used reagents and thermal condition were listed in Table 4.2.

Table 4.2 List of reagents, primer sequences and thermal cycling condition used for multiplex HPg-V2 RT PCR assay.

Reagents	Oligo sequences 5'-3'	Concentration	Used volume (per reaction)	Thermal cycling condition
Reaction mix	NA	2X	12.5µl	60ºC for 30 min, 95ºC for 1 min, followed by 40cycles of 95ºC for 30s and 60ºC for 1 min (read), and 72ºC for 30s, and 37ºC for 40s
Enzyme mix	NA	10U/µl	0.5µl	
HpgV-2 5UTR forward primer	CGCTGATCGTGCAAAGGGATG	10µM	0.4µl	
HpgV-2 5UTR reverse primer	GCTCCACGGACGTCACACTGG	10µM	1.6µl	
HpgV-2 5UTR probe	Cy5-GCACCACTCCGTACAGCCTGAT-BHQ2	10µM	1.6µl	
HpgV-2-NS2 forward primer	GTGGGACACCTCAACCCTGAAG	10µM	0.3µl	
HpgV-2 NS3 reverse primer	CATTGACCGACCTGTCAGGGAAGA	10µM	0.3µl	
HpgV-2 NS2/3 probe	FAM-CCTGGTTTCCAGCTGAGTGCTCC-BHQ1	10µM	0.5µl	

Note: NA: not applicable

2.4. Whole genome sequencing of HPgV-2:

To recover the whole-genome sequences of HPgV-2, HPgV-2 positive samples were re-sequenced using the viral metagenomics assay described above. However, to increase the chance of obtaining more HPgV-2 sequences from each individual samples, the sequencing depth was increased by reducing the total number of samples multiplexed in one run (i.e. from 96 to 6 samples).

To close the remaining gaps (when relevant), several overlapping PCR primers designed based on the obtained HPgV-2 sequences were employed. For each PCR amplification, Super Script III one-step RT PCR with platinum Taq high fidelity DNA polymerase (Invitrogen) was used together with the corresponding PCR primer sets. The obtained PCR products were directly Sanger-sequenced using corresponding PCR primers. PCR primers and PCR conditions are listed in Table 4.3.

Table 4.3 List of primers used to close the gaps of HPgV-2 genomes

Name of primers	Oligo sequences (5'-3')	Thermal cycling conditions
HPgV-2-3491F2	CTTTACTGAGGTCGTGGATG	1 cycle of 30min at 55ºC, 2min at 94ºC and 40 cycles of 15s at 94ºC, 30s at 55ºC and 1 min at 68ºC and 7 min at 68ºC
HPgV-2-4924R2	CAACACCCGCAGTCGATGAC	
HPgV-2-5874F4	GTCTGCTCTGTGCTGGTTGTC	1 cycle of 30min at 60ºC, 2min at 94ºC and 40 cycles of 15s at 94ºC, 30s at 60ºC and 1 min at 68ºC and 7 min at 68ºC
HPgV-2-6811R4	TGCCATTTGTCGCCCCGCCG	
HPgV-2-3491F2	CTTTACTGAGGTCGTGGATG	1 cycle of 30min at 50ºC, 2min at 94ºC and 40 cycles of 15s at 94ºC, 30s at 50ºC and 1 min at 68ºC and 7 min at 68ºC
HPgV-2-R-inAmp2	GAGTATGTTGGTGTCACAGC	
HPgV-2-F-inAmp2	GCTGTGACACCAACATACTC	1 cycle of 30min at 50ºC, 2min at 94ºC and 40 cycles of 15s at 94ºC, 30s at 50ºC and 1 min at 68ºC and 7 min at 68ºC
HPgV-2-4924R2	CAACACCCGCAGTCGATGAC	
HPgV-2-7759F-ref	GAGTCTGTGACGTCAATGGAG	1 cycle of 30min at 55ºC, 2min at 94ºC and 40 cycles of 15s at 94ºC, 30s at 55ºC and 1 min at 68ºC and 7 min at 68ºC
HPgV-2-9284R-3UTR	CATATCAGTCCTGATGGCGCG	
HPgV-2-F2-S62	GAACTAACGCAGCAGCTCTC	1 cycle of 30min at 55ºC, 2min at 94ºC and 40 cycles of 15s at 94ºC, 30s at 55ºC and 1 min at 68ºC and 7 min at 68ºC
HPgV-2-R2-S62	CATAAAAGACCATGGCGCTC	
HPgV-2-F2-S65&278	CTGACCCAATACTCAGTGTG	1 cycle of 30min at 52ºC, 2min at 94ºC and 40 cycles of 15s at 94ºC, 30s at 52ºC and 1 min at 68ºC and 7 min at 68ºC
HPgV-2-R3UTR-S278	GTAAACGCGCGATGTGTCTG	
HPgV-2-1368F1	CCACCAGCACCGATTTCCGC	1 cycle of 30min at 55ºC, 2min at 94ºC and 40 cycles of 15s at 94ºC, 30s at 60ºC and 1 min at 68ºC and 7 min at 68ºC
HPgV-2-R1-1	CATATACATGGCGCACGCTC	
HPgV-2-F1-1-S62	GGCGAGTATTTGGTCTAGAG	1 cycle of 30min at 52ºC, 2min at 94ºC and 40 cycles of 15s at 94ºC, 30s at 55ºC and 1 min at 68ºC and 7 min at 68ºC
HPgV-2-R1-S62	CATCCACGACCTCAGTAAAG	
HPgV-2-F2-2-S65	GTTGAGCTGCTGGAAGAAAC	1 cycle of 30min at 52ºC, 2min at 94ºC and 40 cycles of 15s at 94ºC, 30s at 55ºC and 1 min at 68ºC and 7 min at 68ºC
HPgV-2-R2-2-S65	CTGTAGGATTGCATGCATGG	
HPgV-2-S278-F	CTTGGTATTCGTGCAGTGAG	1 cycle of 30min at 45ºC, 2min at 94ºC and 40 cycles of 15s at 94ºC, 30s at 45ºC and 1 min at 68ºC and 7 min at 68ºC
HPgV-2-S278-R	CTGGAGTAGTTGTGTAGTAC	
HPgV-2-S62-F	GGTATGTCCGAAATCTATGC	1 cycle of 30min at 48ºC, 2min at 94ºC and 40 cycles of 15s at 94ºC, 30s at 50ºC and 1 min at 68ºC and 7 min at 68ºC
HPgV-2-S62-R	GGTACAGTATTTGAGGTAGC	
HPgV-2-S65-F	GTGATCACGGTCATGCATAC	1 cycle of 30min at 52ºC, 2min at 94ºC and 40 cycles of 15s at 94ºC, 30s at 55ºC and 1 min at 68ºC and 7 min at 68ºC
HPgV-2-S65-R	CATAGGTCATATACGCCAAG	

2.5. MiSeq sequence data processing:

Paired-end reads generated by Illumina MiSeq platform, were then processed to remove adapter sequences using Illumina vendor software (Illumina). The resulting sequence data was analyzed by an in-house viral metagenomic analysis pipeline running on a 36-nodes Linux cluster as previously described (190). Subsequently any suspected viral hits were mapped to the corresponding viral genome sequences using Geneious 8.1.5 (Biomatters, San Francisco, CA, US).

2.6. Genetic characterization and Phylogenetic analysis:

Multiple sequence alignment of HPgV-2 sequences available in GenBank and from the present study (Vietnam: n = 5, China: n=2, UK: n = 2 and US: n = 20) and other pegivirus species (n=83) was generated using the ClustalW alignment tool available in Geneious (Biomatters). Maximum likelihood phylogenetic trees based on amino acid sequences of nonstructural regions (NS3 and NS5B) and entire region of coding sequences (CDS) were reconstructed using IQ TREE software (v1.4.3)(246). The LG matrix with empirical amino acid frequencies, a gamma-distribution (4 rates) and invariant sites (LG+F+I+G4) as suggested by IQ TREE was employed to re-construct the phylogenetic trees. Support for individual nodes was assessed using a bootstrap procedure of 10,000 replicates.

Codon-based method (HyPhy) available in MEGA5 (272) was used for measure natural selection pressure on evolution of HPgV-2.

Screening for minor variants was performed using the single nucleotide polymorphism (SNP) tool available in Geneious (Biomatters). A sequencing depth of 300 and minimum variant frequency of >2.5% were chosen as cut-off values.

2.7. Sequence accession numbers:

The HPgV-2 sequences generated in this study were submitted to NCBI GenBank under accession number MH194408-MH194413.

2.8. Ethics

The studies were approved by the corresponding institutional review board of the local hospitals in Viet Nam where patients were enrolled, and the Oxford Tropical Research Ethics Committee. Written informed consent was obtained from either the participant, or the participant's parent or legal guardian.

2.9. Contributions from others:

Clinical samples and data collection were carried out by participating clinicians and research

staff of the respective collaborating hospitals.

3. Results:

3.1. Metagenomic detection of human pegivirus 2 sequences in a serum sample of an adult patient with CA sepsis infection:

As described in Chapter 3, the analysis of metagenomic data revealed that in one serum sample, of 98,344 obtained reads, 5,342, 430 and 273 reads were of HCV, HIV and HPgV-2 sequences, respectively, which was then confirmed by corresponding HCV/HIV/HPgV-2 specific RT-PCRs (Table 3.3). Additional HPgV-2 sequence screening and HPgV-2 RT-PCR testing did not detect HPgV-2 in any of the remaining serum, swab or stool samples of the undiagnosed patients included in metagenomic analysis.

3.2. Prevalence and persistence of HPgV-2 in HCV infected subjects:

In order to explore the prevalence of HPgV-2 in HCV infected patients in Vietnam, a reference-based mapping strategy was applied to screen for HPgV-2 sequences in available viral metagenomic datasets, which were generated from 394 plasma samples of HCV mono-infected patients and 79 plasma samples from patients with HIV/HCV coinfection using the same viral metagenomic procedure employed in the present study (Table 4.1). Subsequently, HPgV-2 genomic sequences were detected in five of 79 HIV/HCV co-infected individuals, but in none of 394 HCV mono-infected subjects from Vietnam (Table 4.1).

To confirm the result of the metagenomic screening, HPgV-2 multiplex RT-PCR was then employed to test the extracted RNA samples from all 79 HCV/HIV co-infected subjects. HPgV-2 RNA was detectable in all five patients whose viral metagenomic results contained HPgV-2 sequences, whilst no additional HPgV-2 was detected in the remaining 74 samples.

Multiplex RT-PCR screening of HPgV-2 RNA in plasma samples of matched controls (78 HIV infected patients and 80 healthy volunteers) of the 79 HCV/HIV co-infected patients revealed

no evidence of HPgV-2. Additionally, HPgV-2 RNA was detected in none of plasma samples from patients with HAV (n=71) and HBV (n=103) infection (Table 4.1).

In order to study the persistence of HPgV-2, the HPgV-2 multiplex RT-PCR was utilized to detect HPgV-2 RNA in follow-up plasma samples collected from six patients (one from the sepsis cohort and five from the HCV trial) whose first samples were HPgV-2 positive (Table 4.4). These six patients were also positive for HIV, HCV and HPgV-2 PCR at enrollment. HPgV-2 RNA was detectable for up to 18 months in three out of five patients who were HCV/HIV co-infected. In one patient, HPgV-2 RNA, but not HCV RNA became undetectable at two follow-up time points (month 6 and 12). HPgV-2 RNA was not detected in the available follow-up serum collected 14 days after enrolment from the sepsis patient.

Table 4.4 Detection of HPgV-2 in longitudinal samples

Patient [#]	Serum samples collected at			
	14 days after enrolment	6 months after enrolment	12 months after enrolment	18 months after enrolment
1	negative*	NA	NA	NA
2	NA	positive	positive	positive
3	NA	positive	positive	positive
4	NA	positive	positive	positive
5	NA	positive*	positive*	negative*
6	NA	positive	negative	negative

Note: *HCV viral load undetectable, NA: not available. [#]All prior collected samples were PCR positive for HCV, HIV and HPgV-2.

3.3. Demographics and clinical characteristics of HPgV-2 infected patients:

Demographic details and clinical presentations of the six HPgV-2 infected patients are presented in Table 4.5. All infected patients were male. The patient with CA sepsis infection was recorded as surviving to 28 days of follow up. All 5 HCV/HIV co-infected patients had CD4 counts >200 cells/µl at baseline and at 6, 12 and 18 months follow up (Table 4.5), but none received specific anti-HCV drugs, which was attributed to drug unavailability/unaffordability during the time of the study period. Only one patient (ID.6) had abnormal alpha-fetoprotein (AFP) levels and fibroscan results (Table 4.5) at baseline and during follow up. During follow up, hepatitis and

splenic abnormalities were detected in 4/5 patients, which were likely attributable to HCV infection (Table 4.5).

Table 4.5 Demographic and clinical features of HPgV-2 infected patients

Patient	Gender	Age (years)	Time point (months)	HCV RNA (+)	HPgV2 RNA (+)	Total Bilirubin (µmol/L)	Direct Bilirubin (µmol/L)	AST (UI/L)	ALT (UI/L)	CD4 count (cells/ µL)	HIV RNA (x10^3 copies/µL)	AFP (mg/ml)	FibroScan result (kPa)	Note
1*	Male	>=18-<40	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2	Male	47	baseline	yes	yes	9.8	0.7	30	24	331	120	1.7	11.8	
			1	NA	NA	4.2	3	91	89	NA	NA	NA	NA	
			2	NA	NA	5.5	3	81	79	NA	NA	NA	NA	
			3	NA	NA	4.7	2.8	44	47	NA	NA	NA	NA	
			6	Yes	yes	4.7	1.6	81	83	518	0.07	2.3	NA	
			9	NA	NA	5.8	2.1	47	67	NA	NA	NA	NA	
			12	yes	yes	6.9	3.4	55	61	364	0.04	2.6	11.8	hepatitis
			15	NA	NA	4.7	2.3	41	43	NA	NA	NA	NA	
			18	yes	Yes	4.8	2.8	37	40	428	undetectable	2.14	6.1	hepatomegaly
3	Male	32	baseline	yes	Yes	4.7	3.4	39	10	288	0.198	0.999	6.5	
			1	NA	NA	13.7	5.9	63	19	NA	NA	NA	NA	
			2	NA	NA	9.1	3.4	33	13	NA	NA	NA	NA	
			3	NA	NA	9	3.8	34	12	NA	NA	NA	NA	
			6	yes	yes	12.8	4.7	50	19	510	0.04	1.68	NA	
			9	NA	NA	6.7	4	48	21	NA	NA	NA	NA	
			12	yes	yes	9.5	5.3	63	25	622	undetectable	1.88	6.2	liver fibrosis, hepatomegaly
			15	NA	NA	7.8	2.4	42	26	NA	NA	NA	NA	
			18	yes	yes	7.6	3.8	42	23	622	undetectable	1.53	7.2	hepatitis
4	Male	35	baseline	yes	yes	7.8	4.9	67	55	290	61.1	2.96	6.4	
			1	NA	NA	6.7	2.4	54	51	NA	NA	NA	NA	
			2	NA	NA	13.2	6.5	66	62	NA	NA	NA	NA	
			3	NA	NA	9.7	3.6	44	52	NA	NA	NA	NA	
			6	yes	yes	10.7	6.3	77	80	411	undetectable	3.1	NA	
			9	NA	NA	8.8	4.6	66	65	NA	NA	NA	NA	
			12	yes	yes	8.8	3.9	76	72	337	undetectable	4	8.5	homogeneous hepatomegaly
			15	NA	NA	9.2	4.3	50	46	NA	NA	NA	NA	
			18	yes	yes	13	6.3	108	129	455	undetectable	4.1	8.1	Splenomegaly, liver fibrosis
5	Male	34	baseline	no	yes	4.3	2.8	33	43	291	70.2	3.67	6.1	
			1	NA	NA	2.6	2.2	28	27	NA	NA	NA	NA	
			2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
			3	NA	NA	9.63	4.15	40.18	32.58	NA	NA	NA	NA	
			6	no	yes	6.5	2.1	35	43	287	undetectable	3.83	NA	
			9	NA	NA	3.1	2.3	68	82	NA	NA	NA	NA	
			12	no	No	5.4	2.6	33	40	484	undetectable	4.48	4.5	
			15	NA	NA	5.7	2.7	33	59	NA	NA	NA	NA	
			18	no	No	6.6	2.6	73	85	546	undetectable	3.9	3	
6	Male	31	baseline	yes	yes	4.5	2.4	52.2	36.5	295	96.8	12.7	22.8	
			1	NA	NA	7.2	1.5	58	42	NA	NA	NA	NA	
			2	NA	NA	14.8	7.2	45	24	NA	NA	NA	NA	
			3	NA	NA	11.5	3.8	44	33	NA	NA	NA	NA	
			6	yes	No	17.1	12.9	64	62	579	undetectable	16.74	NA	
			9	NA	NA	10.2	8.2	114	103	NA	NA	NA	NA	

			12	yes	No	12.3	4.3	114	121	711	undetectable	46.3	26.3	mild liver fibrosis, mild splenomegaly
			15	NA	NA	13.6	7.3	95	91	NA	NA	NA	NA	
			18	yes	No	10.6	4.9	82	89	816	undetectable	61.01	NA	hepatomegaly, splenomegaly

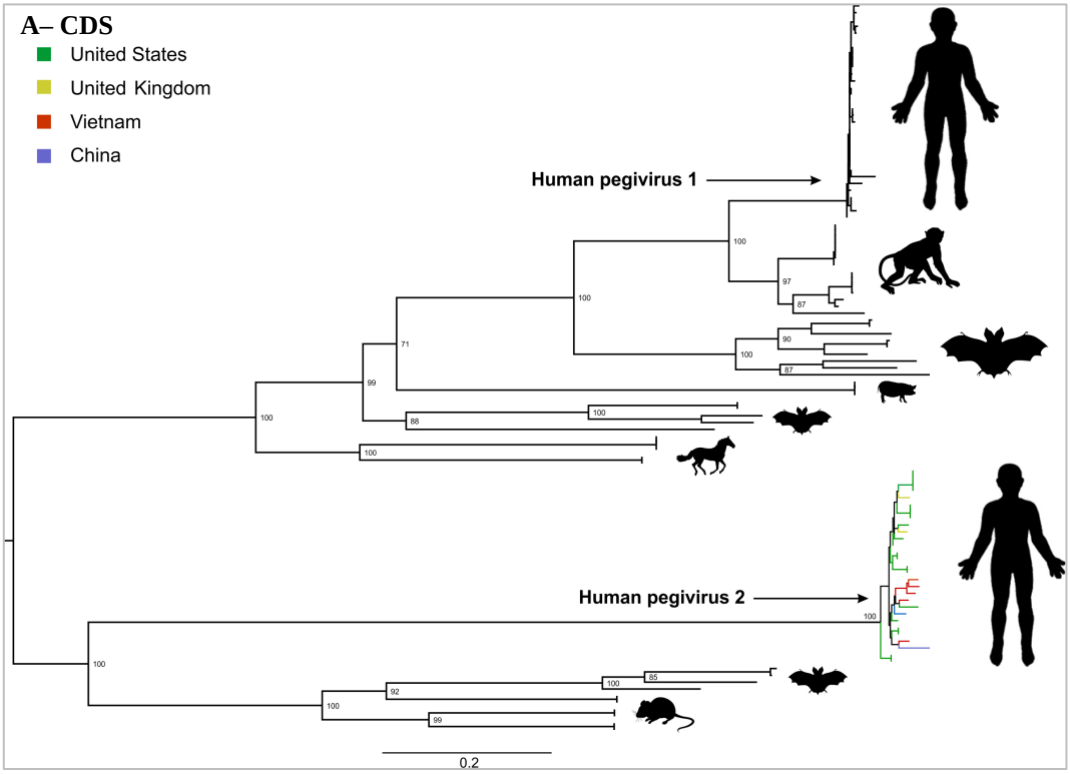
3.4. Genetic characterization of Vietnamese HPgV-2 strains:

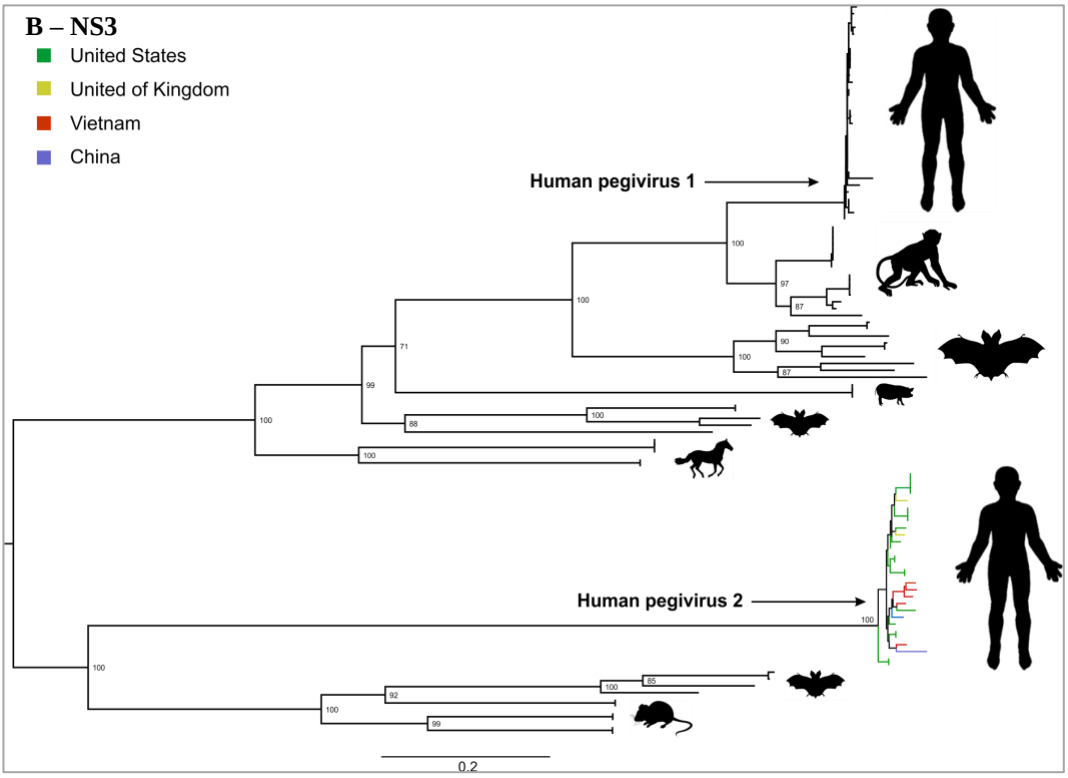
Utilizing a combination of deep sequencing and overlapping PCR approaches, five nearly complete genomes (coverage of >92%) were obtained, whilst in one case a genome coverage of 69.1% was achieved. The results of whole genome sequencing efforts are summarized in Table 4.6.

Table 4.6 Whole genome sequencing results

Patient	Genome coverage % (bp)	Mean of Depth (number of reads per nucleotide)
1	93.3% (8,897 of 9,538)	146 ± 296.6
2	99.0% (9,446 of 9,538)	2060 ± 2453
3	69.1% (6,590 of 9,538)	174 ± 313.3
4	98.0% (9,346 of 9,538)	32531.7 ± 37335.6
5	92.9% (8,857 of 9,538)	34 ± 75
6	95.2% (9,076 of 9,538)	2022.2 ± 3188.3

Pairwise comparison of 5 HPgV-2 polyprotein coding regions obtained in this study showed overall sequence identities at the nucleotide and amino acid level of ≥94.6% and ≥95.3%, respectively. Similar genetic distances were observed while comparing Vietnamese sequences with global sequences as well as between global sequences (Appendix 8), suggesting a close relatedness between Vietnamese HPgV-2 and global HPgV-2 strains. Phylogenetic analyses revealed a tight cluster between Vietnamese viruses and global strains sampled from China, the US and the UK (Figure 4.1).


A– CDS
United States
United Kingdom
Vietnam
China
Human pegivirus 1
Human pegivirus 2
0.2


B – NS3
United States
United of Kingdom
Vietnam
China
Human pegivirus 1
Human pegivirus 2
0.2

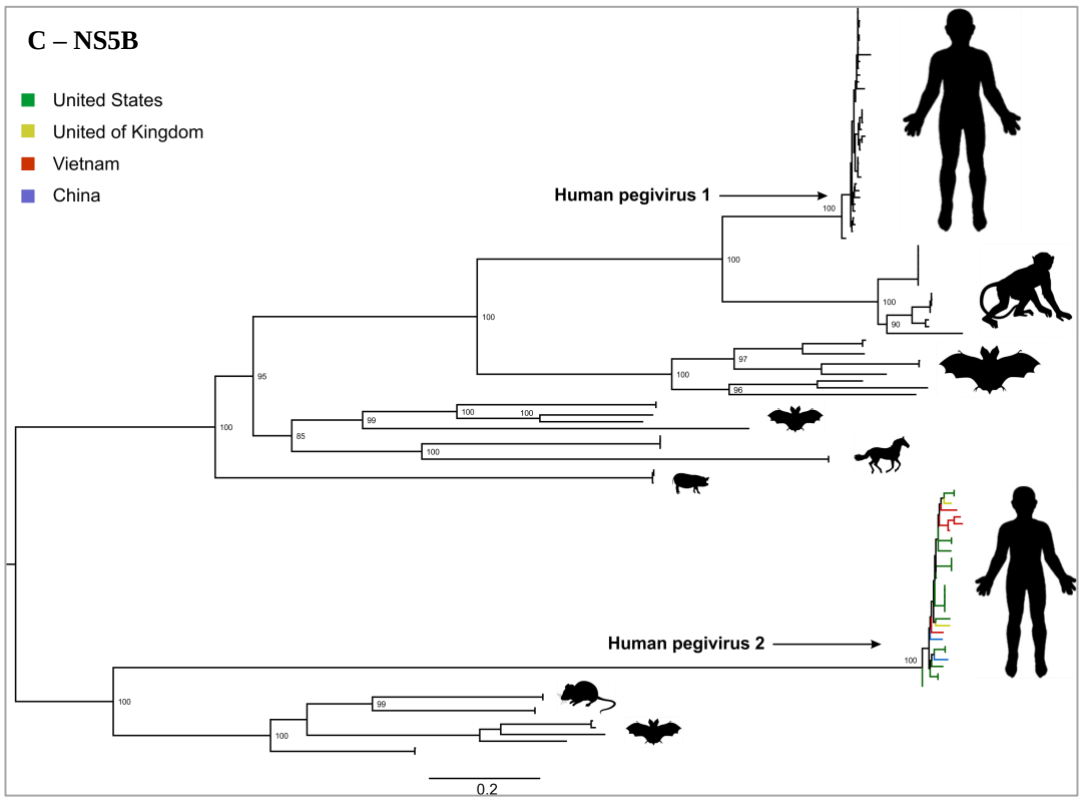


Figure 4.1 Phylogenetic tree of amino acid sequences of global HPgV-2 strains and other pegiviruses. Trees were based on amino acid sequences of CDS (A), NS3 (B) and NS5B (C) proteins, respectively. Vietnamese HPgV-2 strains were denoted in red, and those from the U.S., UK, and China are in green, yellow and blue, respectively. Scale bars indicate number of amino acid substitutions per site.

3.5. Intra-host variants and natural selection:

The depth of deep sequencing results was sufficient for intra-host diversity investigation in 2/6 samples under-investigation. In total, there were 26 and 37 positions carrying minor variations detected in each corresponding dataset, of which 38% (10/26) and 35% (13/37), respectively, were nonsynonymous (Table 4.7).

The estimated ratios of nonsynonymous/synonymous of polyprotein coding and individual protein coding genes of HPgV-2 were well below 1 (Appendix 9), suggesting that the evolution of HPgV-2 was driven by purifying selection.

Table 4.7 List of nonsynonymous minor variations

Patient	Change	Mutation	Site	Gene	Minor variant frequency (%)
2	A -> G	T>A	540	S	2.9
	A -> G	S>G	159	E2	6.2
	T -> C	S>G	1592	E2	2.8
	G -> A	K>R	1786	E2	23.8
	T -> C	L>P	2542	X	3.3
	A -> G	S>G	2664	X	18.5
	C -> T	L>F	5391	NS3	24.8
	A -> C	I>L	5529	NS4A	2.5
	A -> G	R>K	5752	NS4B	12.3
	C -> G	P>A	6690	NS5A	20.7
	C -> T	R>C	8682	NS5B	10.2
	G -> A	R>H	8893	NS5B	6.4
	T -> G	W>G	9294	NS5B	2.6
4	G -> A	N>S	1039	E1	36
	G -> A	T>A	1188	E2	34.5
	G -> A	K>R	1786	E2	19.6
	A -> G	T>A	2271	X	33.3
	C -> T	F>L	2433	X	49.9
	C -> T	I>T	3244	NS2	20.8
	T -> C	P>S	5225	NS3	48
	A -> G	R>K	6304	NS4B	13
	G -> A	A>T	7968	NS5B	28.4
	G -> A	H>R	8386	NS5B	12

4. Discussion:

I report the detection and genetic characterization of HPgV-2 in Vietnam, and describe the observed demographic and clinical characteristics of HPgV-2 infected patients.

Together with previous reports from China, Iran, Cameroon and the U.S. (266–268,270,273,274), my findings further emphasize the strong association between HPgV-2 and HCV, especially HCV/HIV co-infection. The absence of HPgV-2 in 394 HCV infected patients may have been attributed to the small sample size and the fact that the reported prevalence of HIV among HCV-infected patients was ≤6.5% (275,276). Of note, HPgV-2 was detected in only 0.29% of HCV mono-infected patients in China.

Previous reports have shown that HPgV-2 viremia can either be transient or persistent. Likewise, in the present study, HPgV-2 RNA became undetectable after 14 days in a HCV/HIV co-

infected patient presenting with community-acquired infection of unknown origin, but remained detectable in other HCV/HIV patients through up to 18 months of follow-up.

The pathogenic potential of HPgV-2 remains unknown. However, given its low detection rates in blood donors in the U.S. and China (266,268), and absence in healthy subjects (present study), but tight association with HCV/HIV infection, its role in the natural history of HCV/HIV disease/response to treatment warrants further research.

In the era of sequence-based virus discovery, a key question is whether the detected genome represents live virus or a non-replication competent genome. Addressing this question would require recovery of virus in cell culture. However, the detection of minor variations across two HPgV-2 genomes in this study suggests that possibly viral replication had occurred in the infected patients.

Phylogenetically, the close relatedness between Vietnamese HPgV-2 strains and global strains suggests HPgV-2 has a wide geographic distribution.

My study has some limitations. First, I only retrospectively tested available archived samples without formal sample size estimation, which may have explained the absence of HPgV-2 in the remaining 394 HCV patients. Second, a serological assay was not employed to screen for HPgV-2 specific antibodies in patients' plasma. Third, only multiplex PCR with primers based on a limited number of available HPgV-2 sequences were utilized. Therefore, genetically diverse HPgV-2 strains may have been missed. Collectively, the prevalence of HPgV-2 infections in Vietnam may have been underestimated.

Collectively, the results expand our knowledge about geographic distribution, demographics and genetic diversity of HPgV-2. Because HCV and HIV infections are major global public health issues, the extent to which HPgV-2 may interact with HCV/HIV in co-infected patients with possible clinical consequences, warrants further research.

Chapter 5: Viral metagenomic analysis of cerebrospinal fluid from Vietnamese patients with acute central nervous system infections of unknown origin

1. Introduction:

Here I focused on using a mNGS approach to search for known and unknown viruses in CSF samples collected from Vietnamese patients with CNS infections of unknown causes who were enrolled in a hospital-based surveillance study conducted during 2012–2016.

2. Materials and methods:

2.1. The clinical study and selection of cerebrospinal fluid samples for mNGS analysis:

The study used CSF samples collected from patients with suspected CNS infection enrolled in a hospital based surveillance program conducted in Vietnam from December 2012 to October 2016 (277). The study was conducted as part of the VIZIONS (Vietnam Initiative on Zoonotic Infections) project (277), and patient recruitment was carried out at seven provincial hospitals across Vietnam. After collection, as per the study protocol, all CSF samples were tested for a range of pathogens by using the diagnostic work-up of the clinical study (Appendix 3). The remaining volume of the CSF samples were stored at -80°C for further testing.

Here, I focused on samples collected from patients of unknown origin from 4 provincial hospitals in central (Hue and Khanh Hoa), highland (Dak Lak), and southern (Dong Thap) Vietnam (Figure 5.1), representing 3 distinct geographic areas in Vietnam. To increase the chance of detecting a virus in the CSF samples, only patients with CSF leukocyte counts $\geq$5 cells/mm^3 and an illness duration $\leq$5 days were selected.

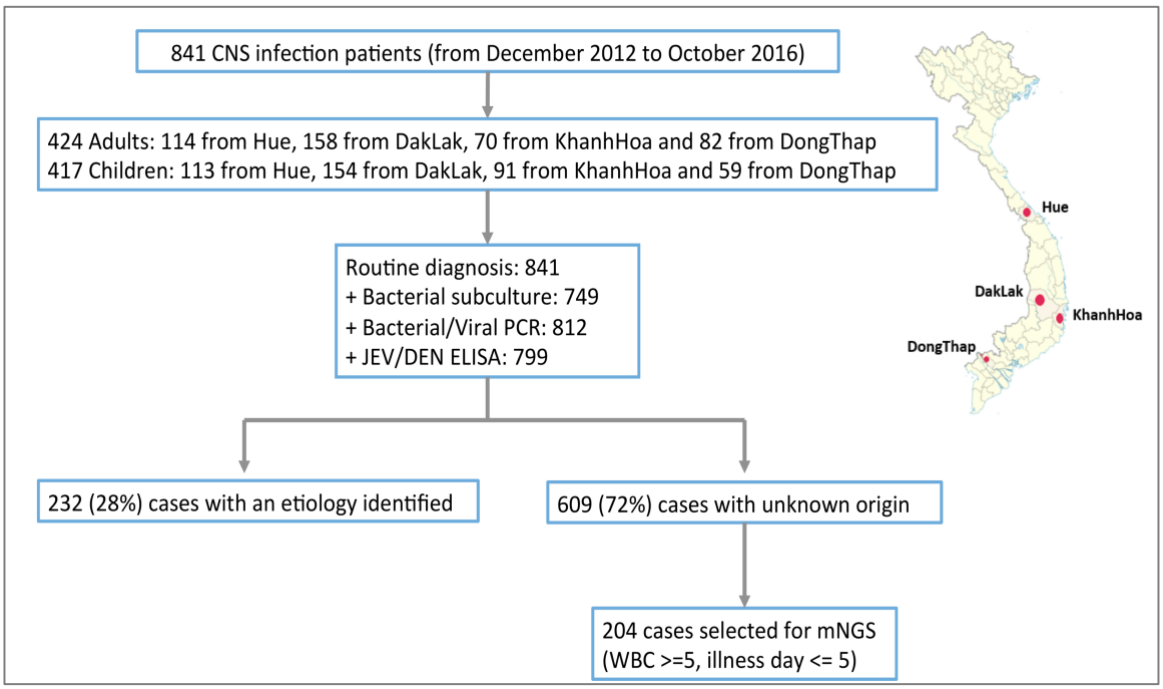


Figure 5.1 Flowchart showing an overview of diagnostic results of the original study of patients with suspected CNS infections admitted to one of the study settings in Vietnam from December 2012 to October 2016. The map indicates places (red dot) where samples were collected.

2.2. mNGS and sequence data analysis:

The optimised mNGS assay described in Chapter 2 was used to analyse the selected CSF samples. To identify potential viral hits, the obtained sequences were analysed using an in-house viral metagenomic pipeline running on a 36-node Linux cluster as described in Chapter 3.

2.3. PCR confirmatory testing of mNGS results:

PCR assays were carried out to confirm mNGS hits for each specific virus identified from the viral metagenomic pipeline. Depending on availability of CSF, the PCR confirmations were performed either on leftover NA or newly extracted NA. A viral mNGS result was considered positive only if it was subsequently confirmed by PCR analysis of the original NA samples. The nucleotide sequences of primers and probes used for PCR confirmatory testing in the current study were presented in Table 5.1 (278).

Table 5.1 List of primers and probes used for PCR confirmatory

Viruses	Oligo sequence (5'-3')			Sources
	Forward	Reverse	Probe	
HBV	GGACCCCTGCTCGTGTTACA	GAGAGAAGTCCACCMCGAGTCTAGA	FAM-TGTTGACAARAATCCTCACAATACCRCAGA-TAMRA	Newly designed
Rotavirus	ACCATCTWCACRTRACCCTC	GGTCACATAACGCCCCTATA	FAM-ATGAGCACAATAGTTAAAAGCTAACACTGT CAA-BHQ1	(231)
Enterovirus	CCCTGAATGCGGCTAAT	ATTGTCACCATAAGCAGCC	CY5-ACCCAAAGTAGTCGGTTCCG - BHQ3	(232)
HIV1	GGTGCGAGAGCGTC	ATGCTRTCATCATYTCTTC		(235)
	ATGGGTRAARGTARTAGAAGAAAAGGG	CTGCCTGRTGYCCYCCCACTA		
Gemycircularvirus	GTGGTAATGGTCGTCGGTATTC	CCTCATCATTCGTAGTAAGCAATCTCA		(243)
	AGTCCTGAATGTTTCCACTCG	CAAGCGTTCCCTCGAAAATGAC		Newly designed
Cyclovirus VN	GAGCGCACATTGAAAGAGCTAAA	TCTCCTCCTTCAATGACAGAAACAAC	FAM-CGADAATAAGGMATACTGCTCTAAAGSTGGCG-BHQ1	(244)
Molluscum contagiosum virus	AACCTACGCTACCTGAAGMTGGA	CAGGCTCTTGATGGTCGARATGGA		(279)

2.4. Serotype identification and phylogenetic analysis:

For EV serotype determination based on the obtained sequences generated by viral mNGS, a publically available genotyping tool was used (280). To determine the relationship between EV strains sequenced in the present study and global strains, I first performed pairwise alignment using ClustalW tool available in Geneious 8.1.5 (Biomatters), and then reconstructed a maximum likelihood phylogenetic tree using IQ tree (v1.4.3) (246). Similar phylogenetic approach was utilized for other viruses.

2.5. Sequence accession number:

The generated sequences of this study were submitted to GenBank (PRJNA561465).

2.6. Ethics:

The study was approved by the corresponding institutional review broad of local hospitals in Vietnam, where the patients were enrolled, and the Oxford Tropical Ethics Committee. Written informed consent was obtained from each study participant or a legal guardian.

2.7. Contributions from others:

Clinical samples and data collection were carried out by participating clinicians and research staff of the respective collaborating hospitals. Although, I led the experiment design and conduct the most of the laboratory work of the laboratory work, I also received support in conducting some of the PCR confirmatory experiments from Ms Le Nguyen Truc Nhu and Ms Nguyen Thi Thu Hong from Emerging Infections group, OUCRU. The in-house viral metagenomic pipeline was conducted at the lab of Prof Eric Delwart at Blood Systems Research Institute, San Francisco, California, United State, with help from Dr Xutao Deng.

3. Results:

3.1. CSF samples available for mNGS analysis:

From the aforementioned clinical study, a total of 841 patients with suspected CNS infection were enrolled from Hue, Khanh Hoa, Dak Lak or Dong Thap provincial hospitals. Of these 609/841 (72%) patients had no etiology identified. The etiological profiles of the patients in whom a pathogen was detected will be separately reported. Of the patients in whom a pathogen was not identified, 204 met the selection criteria and their CSF samples were included for viral mNGS analysis (Figure 5.1).

3.2. Baseline characteristics of included patients:

The baseline characteristics and outcome of the 204 study patients are presented in Table 5.2. Male patients were predominant. A substantial proportion of the patients were seriously ill; fatal outcome was recorded in 22 (11%), while incomplete recovery or deterioration (reflected by being transferred to other hospitals) were recorded in 17% (n=35) and 16.5% (n=34) of the patients, respectively.

Table 5.2 Baseline characteristics and clinical data of patients with acute central nervous system infections enrolled for mNGS analysis of CSF samples, Vietnam, December 2012–October 2016*

Characteristics	Patients with unknown cause enrolled for mNGS (n = 204)	Patients with mNGS negative (n = 174)	Patients with Enterovirus detected (n = 23)	p value†
Sex				
Male	135 (66)	114 (65.5)	15 (65)	
Female	69 (34)	60 (34.5)	8 (35)	
Age, y, median (range)	20.5 (0–92)	24 (0–92)	13 (2–27)	0.005
Location				
Hue	37 (18)	28 (16)	9 (39)	
Dak lak	98 (48)	87 (50)	10 (43.5)	
Khanh Hoa	28 (14)	22 (13)	4 (17.5)	
Dong Thap	41 (20)	37 (21)	0	
3-d fever (at enrollment or preceding 3 d)				
Fever	148 (72.5)	126 (72.4)	17 (74)	0.054
Temperature, C°, median (range)	39 (37.5–42.0)	39 (37.5–42.0)	38.5 (38.0–40.5)	
Fever with unknown temperature	29 (14.2)	22 (12.6)	6 (26)	
No fever	20 (9.8)	19 (11)	0	
Unknown	7 (3.5)	7 (4)	0	
Outcome				
Death or discharge to die	22 (11)	22 (12.6)	0	
Discharge with complete recovery	108 (53)	86 (49.4)	18 (78.3)	
Discharge with incomplete recovery	35 (17)	31 (17.8)	2 (8.7)	
Transfer to another hospital	34 (16.5)	30 (17.2)	3 (13)	
Other (patient request)	3 (1.5)	3 (1.7)	0	
Unknown	2 (1)	2 (1.3)	0	
CSF white cells, cells/mm^3, (median (min-max)	88.5 (5–40,000)	71.5 (5–40,000)	110 (8–1200)	0.343

Note:*Values are no. (%) unless indicated, CSF, cerebrospinal fluid; mNGS, metagenomic next-generation sequencing.
†Statistic comparisons were performed for groups of patients with mNGS-negative results and enterovirus detected, by Mann-Whiney test.

3.3. mNGS results: a general description

Two hundred and four CSF samples were subjected to three NGS runs. A total of 108 million reads (median number of reads per samples (range): 445,412 (430 – 908,890)) were obtained. Of these, viral reads accounted for 0.64% (n=692,731, median number of reads per sample (range): 2,001 (4-268,933)). Excluding common contaminants and commensal viruses such as TTV, which are not reported here, sequences related to a total of eight distinct viral species were identified in 107/204 (52.4%) patients. These are viruses that are either known to be infectious to humans (EVs, rotavirus, molluscum contagiosum virus, human papillomavirus, HIV and HBV), or without evidence of human infections beside prior detection in “sterile” human samples (cyclovirus-VN, gemycircularvirus) (Figure 5.2).

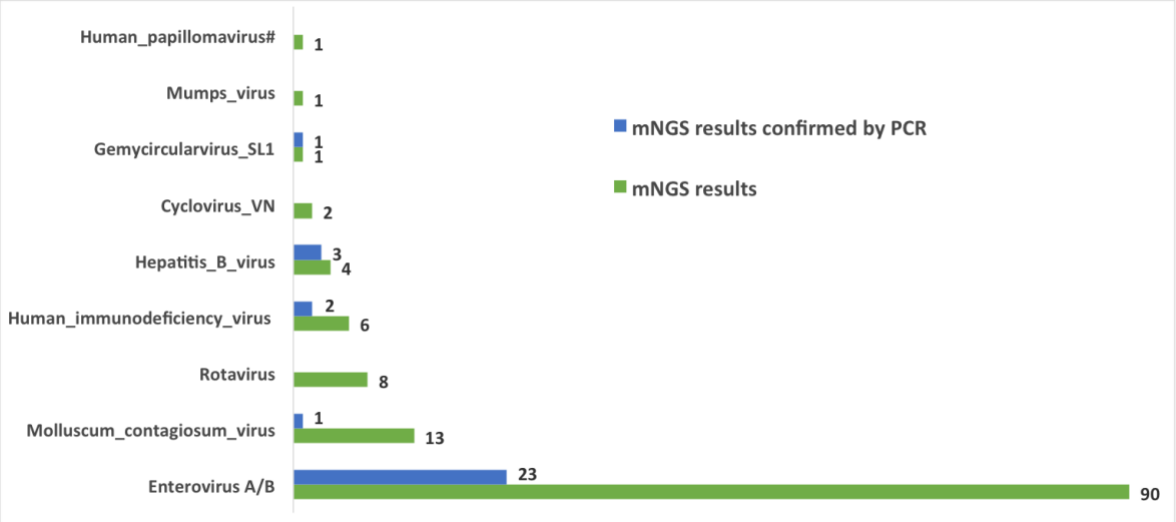


Figure 5.2 Number of cerebrospinal fluid samples with detected viruses by metagenomic next-generation sequencing and then confirmed by virus-specific PCR or reverse-transcription PCR, Vietnam, December 2012–October 2016. Samples were collected from patients with suspected central nervous system infection. Note: #: Confirmatory testing was not performed due to unavailable of PCR assay

3.4. mNGS result assessment by specific PCR analysis:

After viral specific PCR confirmatory testing, the proportion of patients in whom a virus was found by mNGS was reduced from 53% (108/204) to 14.7% (30/204). Accordingly, the number of viral species was reduced from eight to five (Figure 5.2) with EVs being the most common virus detected, accounting for 11.3% (23/204) of the included patients, followed by HBV (n=3), HIV (n=2), gemycircularvirus and MCV (1 each) (Figure 5.2). Because of the focus of the present study and the availability of PCR assays, confirmatory testing for papillomavirus was not carried out.

3.5. Characteristics of the 23 enterovirus infected patients

All 23 EVs infected patients were admitted to hospitals from the central or highland areas (Table 5.3), and none were from Dong Thap province. Male patients were slightly predominant, accounting for 56%. Notably, the EVs infected patients were younger than those who were mNGS negative (Table 5.2). At discharge, incomplete recovery or transfer to other hospitals due to disease deterioration were recorded in 21.7% (Table 5.2).

Temporarily, EVs cases were not detected between January 2015 and December 2016. During 2013 and 2014, there were two main peaks observed from March to July and September to December (Figure 5.3A), with cases from Dak Lak and Khanh Hoa contributing to the first peak (Figure 5.3B&C) and cases from Khanh Hoa and Hue contributing to the second (Figure 5.3C&D).

HBV, HIB, gemycircularvirus and MCV were detected in a total of six patients. Their general baseline characteristics are presented in Table 5.3.

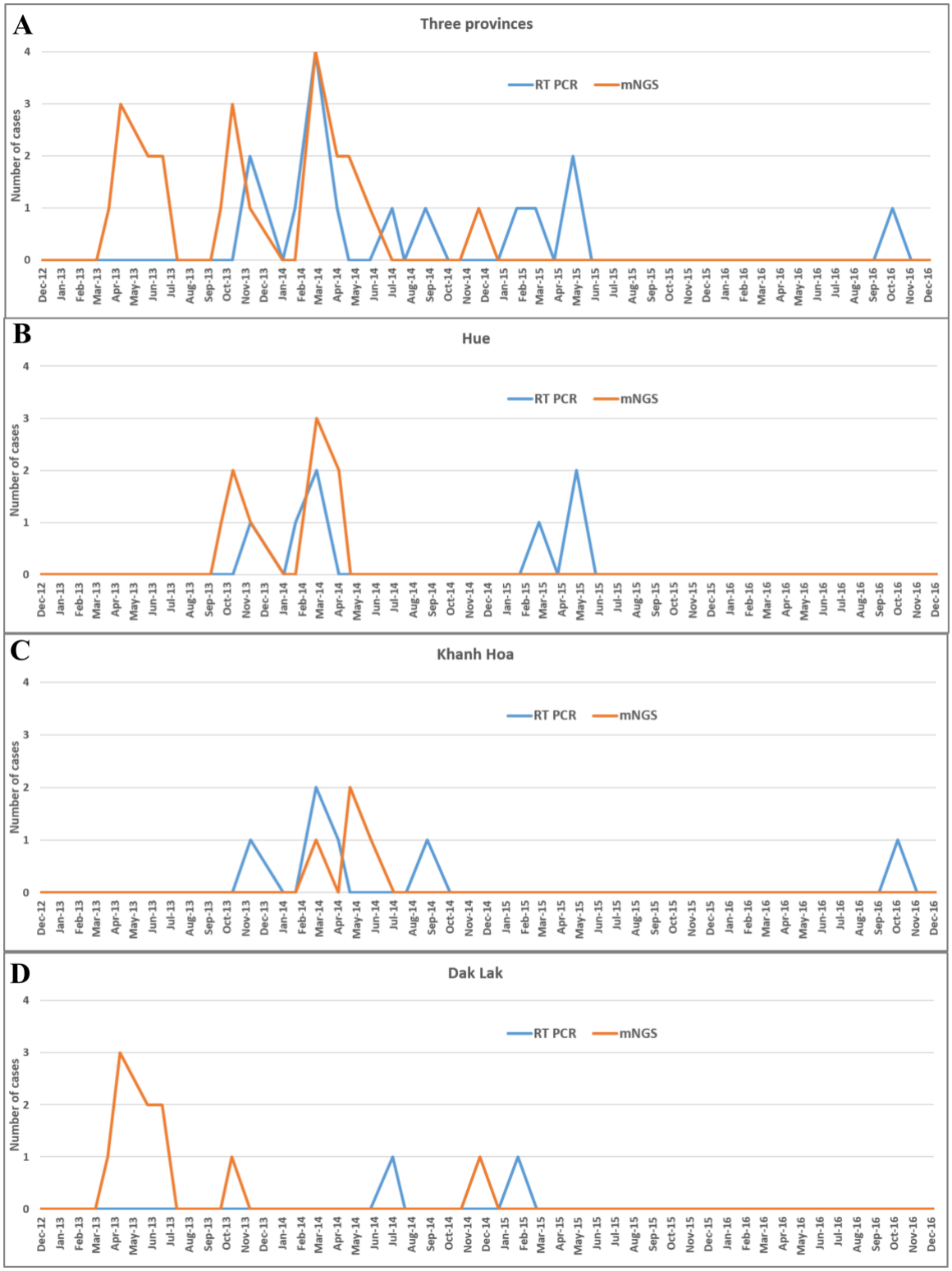


Figure 5.3 Temporal distribution of enterovirus cases detected from CSF samples of patients with suspected CNS infection by mNGS and RT-PCR. Enterovirus RT-PCR results were obtained from the original study.

Table 5.3 Baseline characteristics, and CSF white cell count of patients with viruses detected by mNGS.

Characteristics	Hepatitis B virus (n=3)	Human immunodeficiency virus (n=2)	Molluscum contagiosum virus (n=1)	Gemycircularvirus (n=1)
Male, n(%)	3(100)	2(100)	1(100)	1(100)
Age (median; min-max)	42(16-63)	57.5(49-66)	7	12
Location				
Hue, n(%)	0	0	0	0
Dak lak, n(%)	0	0	1(100)	0
Khanh Hoa, n(%)	2(75)	0	0	0
Dong Thap, n(%)	1(25)	2(100)	0	1(100)
3-day fever (at enrollment or last three days)				
Fever, n(%)	3(100)	1(50)	1(100)	0
Temperature, C°, median (range)	38(38-39)	39	38.5	
Fever with unknown temp, n(%)	0	0	0	1(100)
No fever, n(%)	0	1(50)	0	0
Unknown, n(%)	0	0	0	0
Outcome				
Death or discharged to die, n(%)	0	0	0	0
Discharge with complete recovery, n(%)	2(75)	0	1(100)	1(100)
Discharged with incomplete recovery, n(%)	1(25)	1(50)	0	0
Transferred to another hospital, n(%)	0	1(50)	0	0
Other (patient request), n(%)	0	0	0	0
Unknown, n(%)	0	0	0	0
CSF white cells, cells/mm^3, (median (min-max)	1530 (7-2590)	835(340-1330)	70	60

3.6. Genetic characterization of enteroviruses and gemycircularvirus

mNGS generated sufficient sequence information for EVs serotyping assessment in 11/23 cases. Subsequently, results of serotyping analysis based on the NGS sequences showed that E30 was the most common serotype detected (n=9, 39% of EVs), followed by EV-A71 and enterovirus B80 (1 each, 4.3%). Phylogenetically, the nine E30 strains sequenced in the present study belonged to two distinct genogroups, V and VIIb, and showed close relationship with E30 strains circulating in Russia and elsewhere in Asian countries (including China) (Figure 5.4).

In additional to EV sequences, a gemycircularvirus genome was obtained from a 12 year-old-boy. Phylogenetic analysis revealed that this gemycircularvirus strain was closely related to a gemycircularvirus species previously found in CSF sample from a patient with CNS infections of unknown origin in Sri Lanka (243), with the level of amino acid identities between the two strains being 98.79% for Rep and 99.3% for Cap protein sequences.

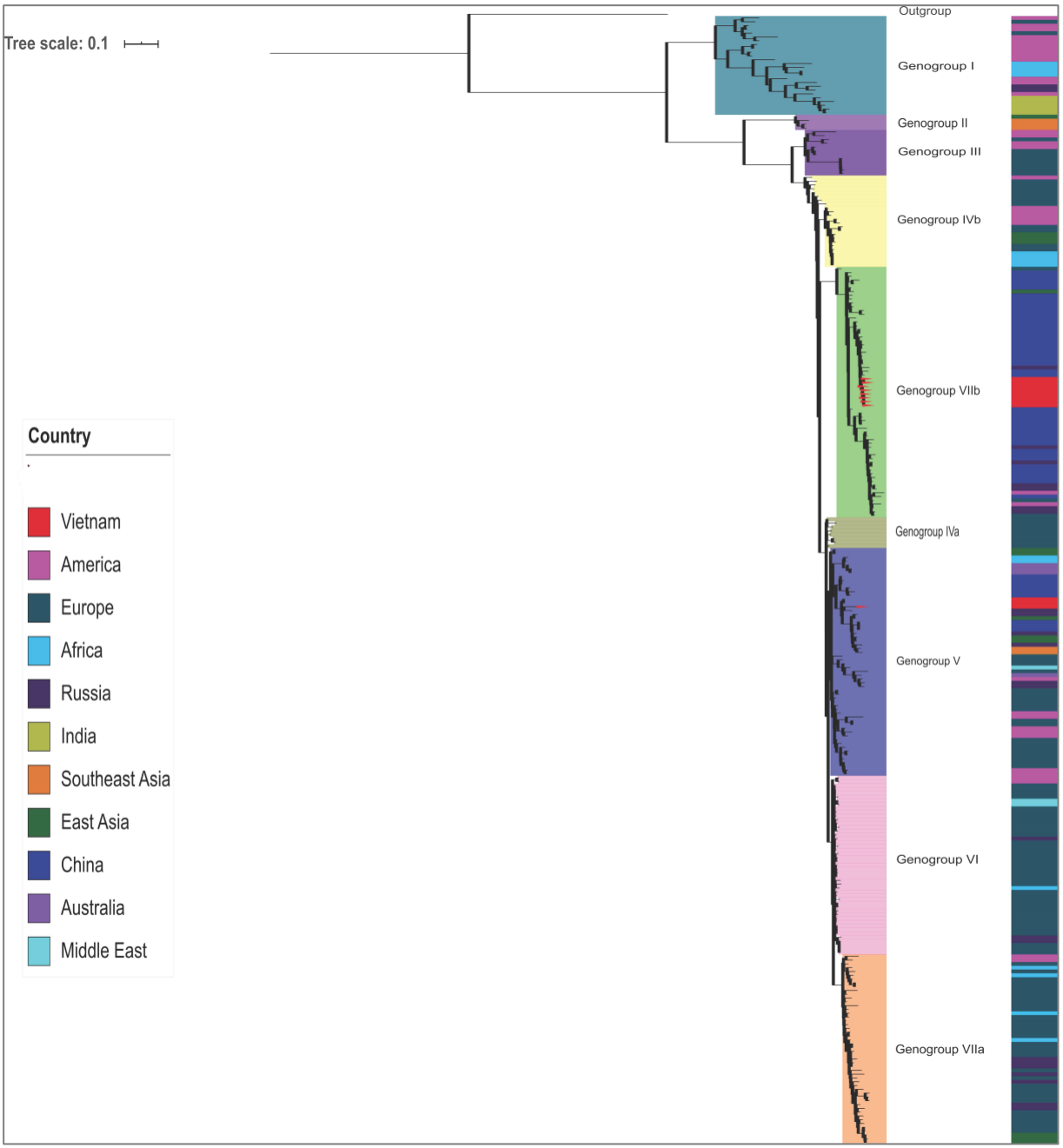


Figure 5.4 Phylogenetic tree of complete vp1 sequences of echovirus 30 (E30) (876 nt) (n = 298) isolated from CSF samples of patients with suspected CNS infection. The inner color-strip indicates for different countries of E30 isolates included in the tree. The outer color-strip indicates 7 genogroups, including genogroup I, II, III, IVa, IVb, V, VI, VIIa and VIIb. The E30 sequences generated by mNGS were highlighted in red color. The outgroup is echovirus 21 Farina.

4. Discussion:

I present a viral mNGS investigation characterizing the human virome in CSF of 204 Vietnamese patients with suspected CNS infection of unknown origin. I successfully detected four human viral pathogens (EVs, HIV, HBV and MCV) and one virus species (gemycircularvirus) of unknown tropism and pathogenicity in a total of 30 of 204 (14.7%) patients. The majority of patients therefore remained without a known etiology. It therefore remains a challenge to identify a plausible viral pathogen in CSF of patients with CNS infections.

EVs were the most common viruses found in 11.3% (23/204) of all analyzed patients (Figure 5.2), the majority being children and young adults. This observation on the age distribution of EVs infected patients was in agreement with observational data from a previous report from Vietnam (95) while the median of age was slightly higher compared with data from other countries (281,282). Geographically, all the EVs infected patients were admitted to hospitals from central and highland Vietnam, and none was from southern Vietnam. The underlying mechanism determining this observed spatial pattern of EVs positive cases in this study remained unknown. It might have been that the sampling time scale was not long enough to capture the circulation of EVs in Dong Thap province. Indeed, EVs were previously reported as one of leading cause of CNS infection in across central and southern Vietnam (91,95,96). Collectively, RT-PCR testing for EVs should therefore be considered in children and young adults presenting with CNS infections.

Of the detected EVs, E30 was the most common serotype. E30 is a well-known pathogen of pediatric aseptic meningitis worldwide (283). Phylogenetically, at global scale E30 belongs to two different lineages with distinct patterns of circulation and spread; one with a global distribution and the other one with geographic restriction within Asia (283). The co-circulation of two E30 lineages in Vietnam suggests that E30 was imported into Vietnam on at least two

occasions. My analyses thus also contributed to the body of knowledge about the genetic diversity of E30 strains circulating in Vietnam.

The detection of blood borne viruses such as HBV and HIV is unlikely to have a direct link with patients' neurological symptoms, although HBV has previously been reported in CSF of patients with CNS infections of unknown origin (284). The detection of HIV in CSF might have been a consequence of traumatic tap occurring during the lumbar puncture as reflected by the presence of a high number of red blood cells in one of two HIV positive CSF (data not shown). However, neuro-invasion of HIV has also been reported (285). Likewise, the pathogenic potential of a gemycircularvirus genome requires further investigation, although the detection of the gemycircularvirus genome in CSF has been reported in several papers in CSF (243,284,286). The detection of MCV and papillomavirus in CSF may result from contamination of viral skin flora during lumbar puncture.

Similar to results of Chapter 3 about discrepancy between mNGS and conventional diagnostic testing (278,284,287), I observed that the majority of mNGS positive results were not confirmed by subsequent viral (RT-) PCRs, especially the sensitive EVs RT-PCR with a limit of detection of around 9 copies per reaction (232). Such results could be a result of bleed over (also called index hopping) of indices from reads of one sample into reads of another co-sequenced on the same Illumina run (288). Notably, the application of double indexes, which was not utilized in the present study, has been shown to significantly reduced although not eliminate the cross-contamination phenomenon between samples in the same run.

My study has some limitations. Firstly, as outlined above, I did not employ double unique index combination strategy per sample as part of the sequencing procedure. The well-known index hoping phenomenon possibly explains the high discrepancy between confirmatory PCR and mNGS results F(53,79,81), and emphasizes the usefulness of dual indexing and including no template controls. As such, I pragmatically chose to verify my mNGS by performing specific

PCR on original materials. Secondly, the DNase treatment step in my assay meant to reduce cellular DNA concentration in CSF, which may reduce the sensitivity of mNGS for the detection of DNA viruses such as HSV (172,289). Thirdly, some of the non-PCR confirmed viral sequences likely originated from contamination of reagents a lingering problem for mNGS (290,291).

In summary, my results emphasize mNGS can detect a broad range of viral nucleic acids in CSF. In spite of extensive investigation, it remains a challenge to establish the etiology in many patients with CNS infections. EVs are important causes of viral CNS infections in Vietnam, and thus should be considered as a differential diagnosis among young patients presenting with CNS infections.

Chapter 6: The utility potential of metagenomics for the diagnosis of central nervous system infections

1. Introduction:

From the diagnostic perspective, a recent prospective study in the United States compared the diagnostic performance of routine diagnostic tests with metagenomic next-generation sequencing (mNGS) and showed that mNGS detected a bacteria or virus in the CSF of 13 of 58 patients presenting with meningoencephalitis that were negative or not assessed for with routine diagnostic tests (287). Otherwise, studies to date were either case reports or retrospectively performed with small sample sizes (175,216), but few have been carried out in limited resource settings like Vietnam. Such studies would have significant implications for both disease surveillance and patient management. Herein, I report the results of a pilot assessing the utility potential of mNGS for the diagnosis of CNS infections.

2. Materials and Methods:

2.1. Setting, patient enrolment and data collection

The study was conducted in a brain infection ward of the Hospital for Tropical Diseases (HTD) in Ho Chi Minh City, Vietnam between January 2015 and September 2016. One of the aims of the study was to improve the diagnosis in patients with meningoencephalitis using mNGS. The study enrolled consecutive adult patients (≥18 years) with an indication for lumbar puncture admitted to the study site during the study period. Patients were excluded if pyogenic bacterial meningitis (cloudy or pus-like CSF) was suspected, lumbar puncture was contra-indicated, or no written informed consent was obtained from the patient or their relatives.

As per the study protocol, CSF samples were collected alongside demographic and clinical data (including discharge outcome), and the results of routine diagnostic testing. After collection, all clinical specimens were stored at -80^{o}C for subsequent analyses, including assessment of mNGS

performance against that of routine diagnostic assays. Here I focused on patients with meningoencephalitis regardless of the results of routine diagnosis. Additionally, as negative controls, I used one CSF from a patient presenting with cerebral hemorrhage and one from a patients with laboratory confirmed anti-N-methyl-D-aspartate receptor (292).

2.2. Routine diagnosis:

As part of routine care at HTD, CSF specimens of patients presenting with brain infections were examined using standard methods when appropriate (Appendix 10). More specifically, all CSF were cultured and/or examined by microscopy for detection of bacterial/fungal/*M. tuberculosis* infection. HSV PCR was carried out on patients presenting with clinically suspected meningoencephalitis. VZV PCR, serological testing for IgM against DENV, JEV or Mump virus was performed if clinically indicated and testing for other pathogens (HSV) was negative (91).

2.3. mNGS assay:

A recent report showed that the DNase treatment step utilized as part of mNGS procedure to reduce the host DNA could hinder the detection of HSV in CSF samples (293). This is because CSF is likely to contained naked DNA of HSV rather mature viral particles. Therefore, to allow for the detection of both RNA and DNA viruses, each CSF sample was subjected to two different metagenomic approaches, namely RNA-virus and viral DNA-virus workflows (Figure 6.1). For RNA-virus workflow, the procedure of mNGS assay optimized in Chapter 2 with sample pre-treatment approach #1 (simultaneous DNase and RNase treatment of the original sample without pre-centrifugation) was used. For viral DNA-virus workflow, viral DNA was directly isolated from 200µl of CSF samples without the nuclease treatment step using DNeasy blood and tissue kit (QIAgen GmbH) and was recovered in 50ul of elution buffer. Both random PCR products and isolated viral DNA were separately sequenced on an Illumina MiSeq platform. The sequencing procedure was described in Chapter 2.

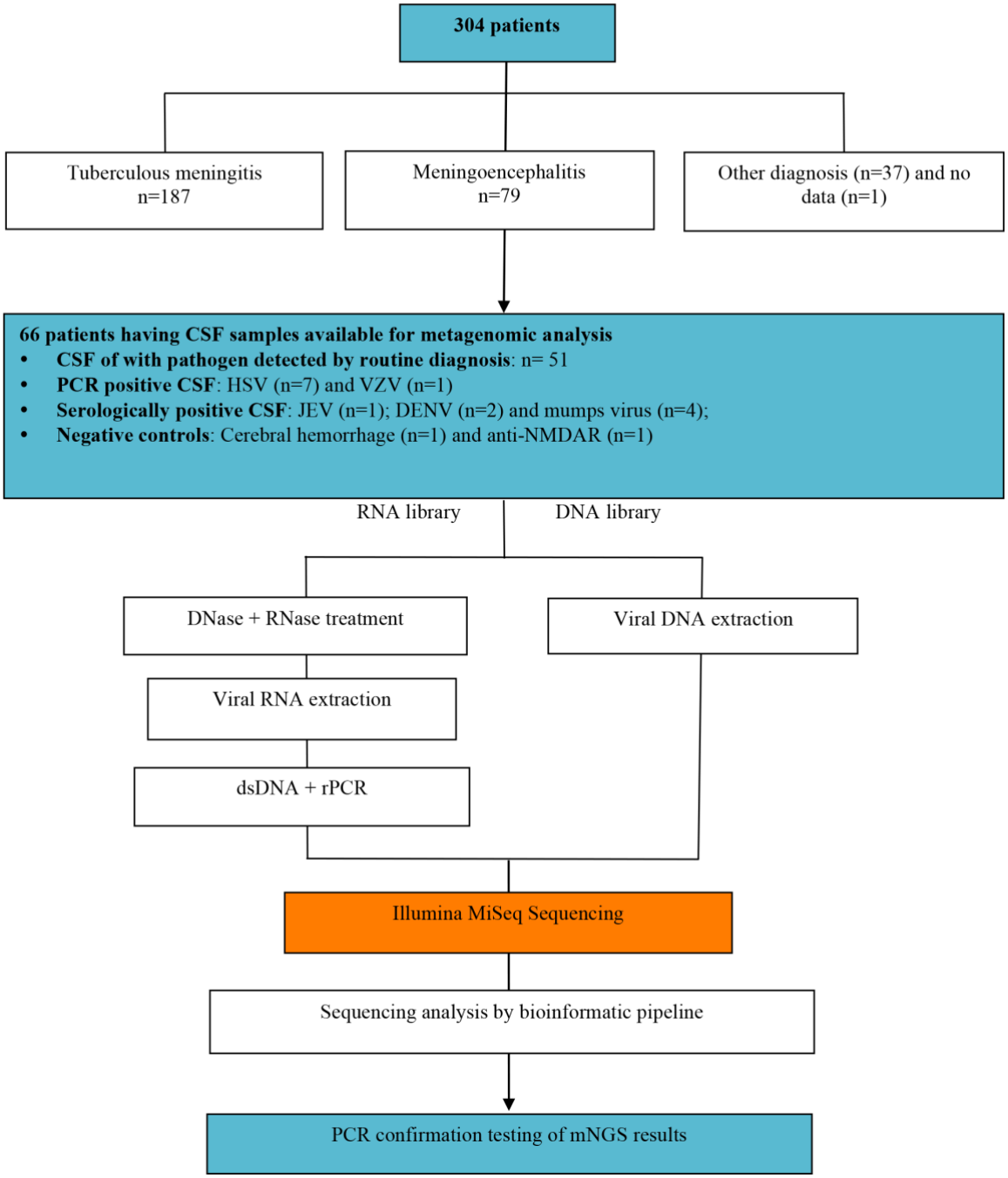


Figure 6.1 Flowchart illustrating an overview about the DNA and RNA virus workflows. Abbreviations: CSF, cerebrospinal fluid; DENV, dengue virus; ds, double-stranded; JEV, Japanese encephalitis virus; mNGS, metagenomic next-generation sequencing; PCR, polymerase chain reaction

2.4. mNGS data analysis:

The mNGS data generated by Illumina MiSeq platform was analysed using an in-house viral metagenomic pipeline running on a 36-nodes Linux cluster as described in Chapter 3.

2.5. PCR confirmation of viral hits detected by mNGS and expanded PCR testing:

Because of the uncertainty in the diagnostic performance of mNGS, and the focus of the present study, specific PCRs were performed to confirm mNGS hits matching to the genomes of neurotropic viruses. Viruses of unknown neurotropic property and well-known contaminants of mNGS dataset were not pursued further by subsequent PCR analysis. The PCR experiments for neurotropic viruses were either carried out on leftover extracted RNA/DNA after the mNGS library preparation experiments or on newly extracted NA. An mNGS result was only considered positive if it was subsequently confirmed by a corresponding viral PCR analysis of original NA materials derived from corresponding individual samples. All PCR primers and probes used were derived from previous publications (231,232,294) (Appendix 11).

2.6. GenBank accession numbers:

Metagenomics data were deposited to NCBI (GenBank) under the SRA accession: PRJNA58865.

2.7. Ethics:

The clinical study received approvals from the Institutional Review Board of the HTD and the Oxford Tropical Research Ethics Committee of the University of Oxford. Written informed consent was obtained from each study participant or relative (if the patient was unconsciousness).

2.8. Contributions from others:

Clinical samples and data collection were carried out by participating clinicians and research staff of the respective collaborating hospitals. Although, I led the experiment design and conduct the most of the laboratory work of the laboratory work, I also received support in conducting

some of the PCR confirmatory experiments from Ms Le Nguyen Truc Nhu and Ms Nguyen Thi Thu Hong from Emerging Infections group, OUCRU. The in-house viral metagenomic pipeline was conducted at the lab of Prof Eric Delwart at Blood Systems Research Institute, San Francisco, California, United State, with help from Dr Xutao Deng.

3. Results:

3.1. Baseline characteristics of the patients included for mNGS

During the study period, a total of 304 patients were enrolled in the clinical study, including tuberculous meningitis (n=187), meningoencephalitis (n=79), other diagnosis (n=37) and no data (n=1). The results of the tuberculous meningitis diagnostic arm have been published elsewhere (295). Of the 79 patients with a discharge diagnosis of meningoencephalitis, 66 (84%) had CSF samples available for mNGS analysis (Figure 6.1). These patients were the focus of the present study regardless of the results of routine diagnosis.

The baseline characteristics of the 66 patients included in the study are presented in Table 6.1. HIV testing was carried out on 24 patients but none was positive. Male patients were predominant. On admission, 35% of the patients were comatose (Glasgow Coma Score <13). Routine diagnostic tests identified a virus in 15/66 (22.7%) the patients (Figure 6.2 and Appendix 12), with HSV being the commonest cause (n=7), followed by Mump virus (n=4), DENV (n=2), JEV (n=1) and VZV (n=1) (Figure 6.2). One patient died, and almost all (n=58) has some neurological deficit at discharge from hospital (Table 6.1).

Table 6.1 Baseline characteristics of the study patients and patients infected with herpes simplex virus, enterovirus or mumps virus

Characteristics	Total (n=66)*	HSV(n=7)**	EVs (n=7)[#]	Mumps virus (n=5)[$]
Demographics				
Gender (male), n(%)	39 (59)	4 (57)	5/7 (71)	5 (100)
Age in years	35 (15-84)	45 (25-53)	32 (22-57)	39 (32-61)
Illness day on admission (days)	5 (1-30)	5 (2-14)	3.5 (2-6)	3 (2-5)
Duration of hospital stay (days)	5 (1-76)	5 (3-67)	2 (1-4)	4 (3-35)
HIV status, n (%)				
Positive	0	0	0	0
Negative	24 (36)	1 (14)	4 (57)	1 (20)
Unknown	42 (64)	6 (86)	3 (43)	4 (80)
Clinical signs and symptoms, n(%)				
Fever	58 (88)	7/7 (100)	6 /7 (86)	5 (100)
Headache	58 (88)	7/7 (100)	6 /7 (86)	5 (100)
Irritability	15 (23)	1/7 (14)	1/7 (14)	0
Lethargy	18 (28)	3/6 (50)	1/7 (14)	0
Vomiting	34 (52)	4/6 (67)	5/7 (71)	3 (60)
Seizures	23 (36)	2/6 (33)	0/7	2 (40)
Conscious	46 (70)	6/7 (86)	1/7 (14)	2 (40)
Skin rash	6 (9)	0/7	0/7	0
Hemiplegia	5 (8)	2/7 (29)	0/7	0
Paraplegia	1 (2)	0/7	1/7 (14)	0
Tetraplegia	1 (2)	0/6	0/7	0
Neck stiffness	45 (68)	6/7 (86)	5/7 (71)	3 (60)
Glasgow coma score of ≤8	7 (11)	3/7 (43)	0/7	1 (20)
Glasgow coma score of 9-12	16 (24)	2/7 (29)	1/7 (14)	1 (20)
Glasgow coma score of 13-15	43 (65)	2/7 (29)	6 /7 (86)	3 (60)
CSF cells and biochemistry				
White cells (cells/µL)	101 (0-4183)	708 (38-1571)	503 (20-961)	683 (27-2146)
Neutrophils (%)	13 (0-96)	9 (2-61)	24 (0-47)	18 (3-23)
Lymphocytes (%)	86.5 (1-100)	91 (39-98)	76 (53-99.9)	82 (77-97)
Protein (g/L)	0.7 (0.2-8.9)	1.36 (0.75-2.17)	0.71 (0.47-1.18)	0.67 (0.45-2.42)
CSF/Blood glucose ratio	0.61 (0.34-1.04)	0.55 (0.47-0.61)	0.71 (0.59-0.85)	0.52 (0.49-0.81)
Lactate (mmol/L)	2.65 (1.4-14.03)	3.52 (2.02-4.83)	2.5 (1.9-3.8)	2.9 (1.9-4.3)
Antiviral treatment, n(%)				
Oral acyclovir	2 (3)	NA	NA	NA
Intravenous acyclovir	8 (13)	6/6 (100)	NA	NA
Oral valacyclovir	44 (72)	NA	NA	1 (20)
Modified Rankin Scale at discharge[&], n (%)				
0	8 (13)	1/7 (14)	1/7 (14)	1 (20)
1	12 (19)	0	1/7 (14)	3 (60)
2	10 (15)	0	4/7 (58)	0
3	25 (39)	3/7 (43)	1/7 (14)	1 (20)
4	4 (6)	0	0	0
5	4 (6)	3/7 (43)	0	0
6	1 (2)	0	0	0

Note: Continuous variables are presented as medina (range); *Denominators may vary slightly, **diagnosed by current standard tests of the routine diagnosis, [#]diagnosed by mNGS followed by PCR confirmatory testing; [$]diagnosed by current standard tests, expanded PCR testing and mNGS combined; [&]0: full recovery with no symptoms, 1: No significant disability, 2: Slight disability, 3: Moderate disability, 4: Moderately severe disability, 5: Severe disability, and 6: Dead.

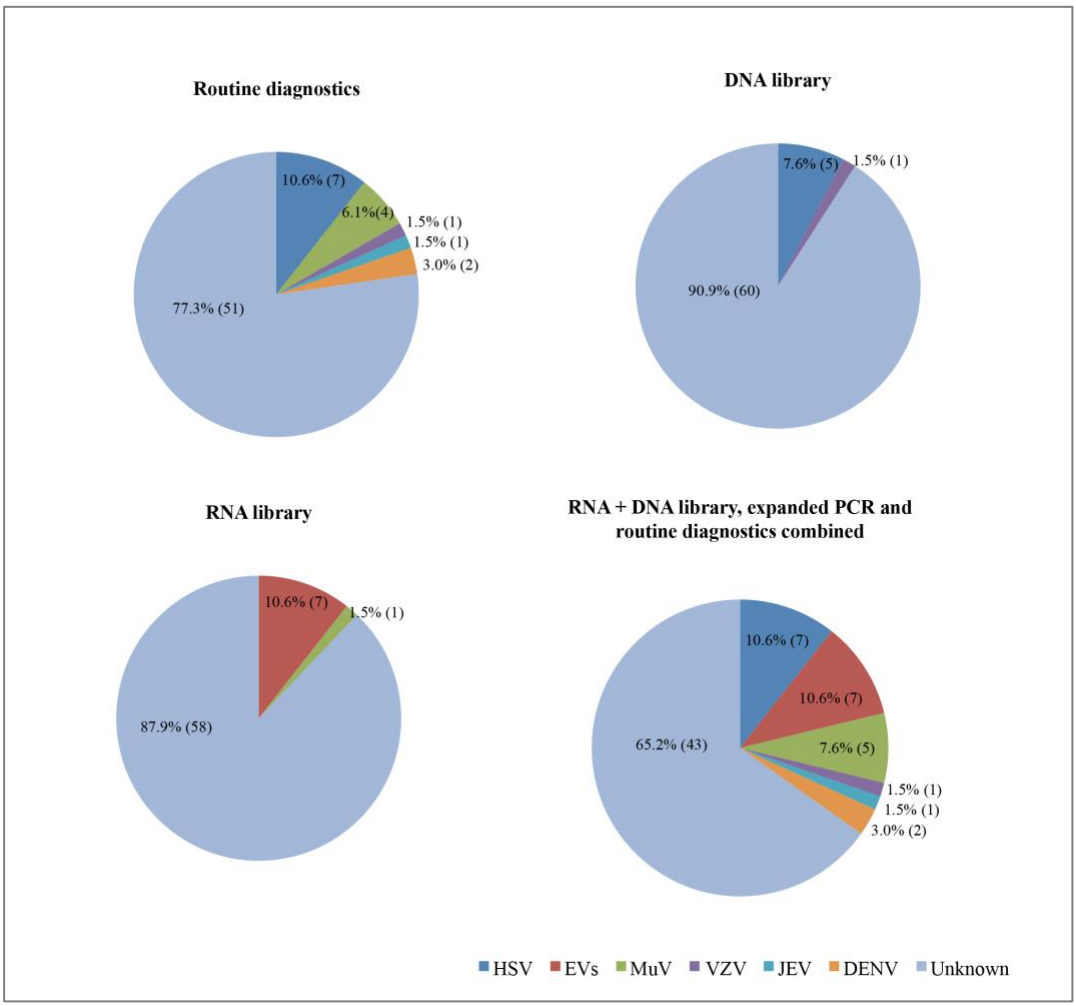


Figure 6.2 Results of mNGS investigations using DNA/RNA workflows and routine diagnostics as well as expanded PCR testing. Abbreviations: DENV, dengue virus; EV, enterovirus; HSV, herpes simplex virus; JEV, Japanese encephalitis virus; MuV, mumps virus; PCR, polymerase chain reaction; VZV, varicella zoster virus.

3.2. An overview of mNGS:

The 68 included CSF samples (including two negative controls) were separately sequenced using both DNA- and RNA-virus workflows in a blinded fashion. Subsequently, a total of 62,565,802 and 49,233,869 reads were obtained from the DNA and RNA libraries, respectively (Appendix 13). Sequences related to 29 viral species were detected, with 23 found in the RNA and seven found in the DNA library (Figure 6.2 and Figure 6.3). The detected viruses included viruses known to cause CNS infections, and those with unknown neurotropic property (TTV (n=14) and herpes virus 8 (n=4)). Additionally, previously reported common contaminants of

mNGS dataset were also found (290,296), almost exclusively found in the RNA-virus library (Figure 6.3).

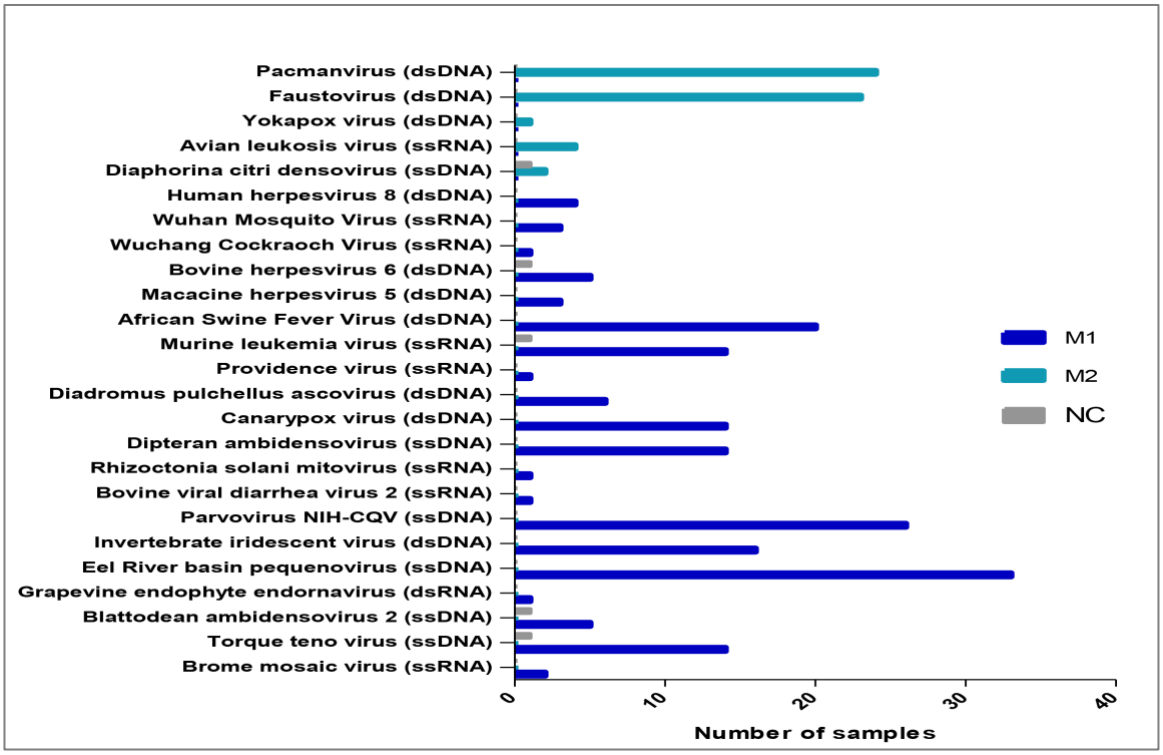


Figure 6.3 Bar chart showing the frequency of common contaminants and viruses of unknown neurotropic property (human herpes virus 8 and Torque teno virus) found in cerebrospinal fluid (CSF) samples by both DNA (M2) and RNA (M1) workflows and viruses in negative control (NC) CSF. Abbreviations: ds, double-stranded; ss, single-stranded.

3.3. Detection of viruses in CSF samples positive by routine diagnosis

Of the 15 CSF samples positive either by PCR or serological testing as part of routine care, mNGS was able to detect a viral pathogen in 5/7 HSV, 1/1 VZV, 1/4 Mump virus, 0/2 DENV and 0/1 JEV positive samples (Figure 6.2). None of the HSV and VZV sequences were found in the library of the RNA-virus workflow (Table 6.2).

Table 6.2 Results of viral PCR and metagenomic analysis

CSF number	Virus	Real-time PCR Ct value	Detected by PCR as part of routine care (Y/N)	Total metagenomic reads	Number of unique viral reads	(%) of viral reads*	mNGS library
1	HSV	25.01	Y	326,396	49	0.015	DNA
2	HSV	28.01	Y	588,504	184	0.031	DNA
3	HSV	30.36	Y	996,348	6	0.001	DNA
4	HSV	23.77	Y	1,145,710	243	0.021	DNA
5	HSV	28.71	Y	346,166	11	0.003	DNA
6	HSV	unavailable	Y	1,345,954	0	0.000	DNA
7	HSV	31	Y	891,566	0	0.000	DNA
8	VZV	22.7	Y	1,335,288	152	0.011	DNA
9	Mumps	35.2	ND	975,714	6	0.001	RNA
10	Enterovirus	33.36	ND	539,752	21	0.004	RNA
11	Enterovirus	34.25	ND	635,310	38	0.006	RNA
12	Enterovirus	34.79	ND	765,564	10152	1.326	RNA
13	Enterovirus	34.78	ND	732,634	89	0.012	RNA
14	Enterovirus	31.23	ND	988,668	2415	0.244	RNA
15	Enterovirus	32.3	ND	594,964	100	0.017	RNA
16	Enterovirus	35.65	ND	543,912	21	0.004	RNA
17	Enterovirus	Negative	ND	579,486	2	0.000	RNA
18	Enterovirus	Negative	ND	571,902	2	0.000	RNA
19	Enterovirus	Negative	ND	720,042	4	0.001	RNA
20	Enterovirus	Negative	ND	511,608	1	0.000	RNA
21	Enterovirus	Negative	ND	818,654	2	0.000	RNA
22	Enterovirus	Negative	ND	513,428	5	0.001	RNA
23	Enterovirus	Negative	ND	1,197,290	13	0.001	RNA
24	Enterovirus	Negative	ND	923,908	4	0.000	RNA
25	Enterovirus	Negative	ND	993,918	1	0.000	RNA
26	Enterovirus	Negative	ND	1,302,784	20	0.002	RNA
27	Enterovirus	Negative	ND	1,628,722	7	0.000	RNA
28	Enterovirus	Negative	ND	1,181,716	24	0.002	RNA
29	Enterovirus	Negative	ND	926,462	22	0.002	RNA
30	Enterovirus	Negative	ND	938,524	20	0.002	RNA
31	Enterovirus	Negative	ND	1,028,194	12	0.001	RNA
32	Enterovirus	Negative	ND	1239458	4	0.000	RNA
33	Rotavirus	Negative	ND	1176486	24	0.002	RNA

Note: *denominators are the total reads of the corresponding samples, ND: not done, Y:yes, N:no

3.4. Detection of sequences related to human pathogenic viruses in CSF negative by routine diagnosis and results of PCR assessment of mNGS results

Of the 51 CSF samples that were negative by routine diagnosis, sequences related to neurotropic viruses were found in 24 (48%) by mNGS (Table 6.2). The detected viruses included EVs (n=23) and rotavirus (n=1). Additionally, of the two CSF samples from non-CNS affected patients, one had four sequences related to enterovirus detected by mNGS.

After PCR confirmation testing of CSF samples in which a viral hit was detected by mNGS, the rotavirus case and the negative control CSF, in which EVs related sequences were detected became negative (Table 6.2). The number of EVs positive CSF was reduced from 23 to seven,

with more enteroviral sequences being recorded in PCR confirmed than in the un-confirmed group (Table 6.2). Of these, three had genome coverage of 61%, 78% and 90%, including one echovirus 6 and two E30. Notably, the majority (12/16, 75%) of EV PCR negative samples had EV reads identical to those obtained from samples with a high abundance of EV sequences (including samples #12 and #14), with which they shared an index (Appendix 14). The data this suggested the potential of barcode bleed through during the sequencing procedure.

3.5. Results of expanded PCR testing and sensitivity assessment of mNGS using PCRs as reference assays:

Because PCR testing for viruses (EVs and Mump virus) was not performed as part of routine diagnosis, to further assess the prevalence of these viruses in the study patients, PCR testing was expanded to CSF samples that were negative by mNGS analysis. Subsequently, only Mump virus was detected by PCR in three CSF samples, including two positive by serological testing as part of standard care (Ct values: 36 and 40), and one negative by mNGS (Ct value: 40). Serological testing for Mump virus in this patient was not done as part of routine care. Thus a combination of serology and molecular assays (PCR and mNGS) increased the diagnostic yield from 22.7% (15/66) to 34.8% (23/66) (Figure 6.2).

mNGS identified a viral pathogen in 14/19 CSF samples that were positive by PCR analysis (including routine diagnosis and expanded testing). Additionally, mNGS detected EVs in 16/47 CSF samples that were negative by PCR in subsequent analysis. Using PCRs as reference assays, the sensitivity and specificity of mNGS were 74% (14/19) and 66% (31/47), respectively. Of the PCR positive samples, there was no difference in the leukocyte counts between the mNGS negative and positive groups (median (range): 331 (27-2146) vs. 356 (22-4183), p=0.82)).

4. Discussion:

I report the results of an investigation assessing the utility of mNGS approach for the diagnosis of viral etiology in the CSF of 66 consecutively treated patients with meningoencephalitis. The patients were admitted to a tertiary referral hospital in Ho Chi Minh City, Vietnam, and the majority (51%) had moderate/severe disability at discharge. The results showed that in a single test, metagenomics could accurately detect NA of a wide range of neurotropic viruses in CSF of the 66 participants, whose diagnoses were only established by extensive PCR testing targeted at a broad range of pathogens. Notably, of these 66 patients, seven (11%) EVs infected patients were initially left undiagnosed at hospital discharge because physicians did not consider EVs diagnosis as part of routine care. EVs infection should therefore be considered as an important differential diagnosis in adults presenting with meningoencephalitis (97), and should be excluded (e.g. by PCR testing) prior to mNGS analysis.

Although antivirals are currently not available for most encephalitis causing viruses, rapid and accurate detection of viral etiology in patient samples remains critical to inform clinical management, such as avoiding unnecessary antibiotic prescription, and public health policy makers. Thus, testing for a wide spectrum of pathogens is essential to maximize the diagnostic yield in patients presenting with meningoencephalitis. Under this circumstance, single pan-pathogen assay such as mNGS is a useful approach, given the limited amount of CSF samples and resources available for microbial investigation, especially in low- and middle-income countries like Vietnam. However, the failure of mNGS to detect nucleic acids of JEV and DENV in serologically positive CSF samples emphasize that testing for pathogen specific antibodies remains an important diagnostic pathway in patients presenting with meningoencephalitis as viral nucleic acids of some viruses (e.g. flaviviruses) may not be present in the collected CSF.

The sensitivity of my mNGS workflows is comparable with that of recent mNGS studies (214,287). Low viral load may be a factor explaining the failure of mNGS to detect HSV and

Mump virus in CSF samples with real time PCR Ct values of 31 (HSV), and 36, 40 and 40 (Mump virus). Because viral reads only accounted for a small proportion of total mNGS reads, increasing the sequencing depth per samples would likely increase the sensitivity of mNGS. However, this increases the sequencing costs. The failure of the RNA workflow, which incorporated the DNase digestion step to detect HSV VZV in all the tested CSF samples, supported the finding from a recent report regarding the impact the nuclease treatment step prior to nucleic acid isolation for metagenomics (172).

Currently, there have been no established robust criteria that can reliably define a true mNGS positive without the requirement of confirmatory testing. Criteria such as the presence of at least three reads mapped to three different genomic region of a virus genome or the absence of viral reads in negative controls have recently been proposed (175,287,297). Such approaches are hindered by the well-known cross-talk contamination phenomenon, occurring as part of mNGS procedure (297), which however can be dramatically reduced through the use of dual barcoding strategy recently developed (264). Because the dual barcoding strategy was not employed, cross-talk contamination may explain for the obtained specificity of 66%, which is lower than the reported data from a previous study (214). Alternatively, the low specificity may have been attributed to the degradation of stored viral RNA and/or the low abundance of viral RNA in the tested samples, leading to the failure of EV PCR to replicate some of the mNGS findings. Retrospectively, the specificity of mNGS would have increased to 83% if a threshold of six reads or above was considered as positive (Table 6.2), suggesting a correlation between the number of mNGS reads and PCR confirmatory results. Collectively, the specificity of mNGS based diagnostic approach could potentially be improved through the use of proper barcoding strategy and/or criteria such as those based on the number of unique viral reads obtained from a sample under investigation, which merit further research.

Similar to previous reports (290,296), numerous common contaminants of mNGS dataset (e.g. parvovirus, densovirus) were found in both DNA- and RNA virus libraries in our study. Although it is likely that those contaminants are derived from laboratory reagents (e.g. extraction kits) (290), their potential impacts on the performance of mNGS, especially in terms of sensitivity and specificity, remains unknown.

The strengths of my study include that it was conducted on consecutive cases, minimizing selection bias. CSF samples were analyzed individually, and mNGS hits were re-confirmed by specific PCR, allowing for back-to-back comparison between mNGS and viral PCR. However, my study has some limitations. First, it was conducted on stored CSF samples. Second, I only focused on viruses, while meningoencephalitis can be caused by non-viral agents such as intracellular bacteria (rickettsiae) (298). Third, I did not test other clinical samples. Of note, JEV has recently been detected in urine of patients presenting with meningoencephalitis (299,300). Last but not least, the inclusion no template controls in addition to the two non-infectious CSF samples would have better captured the spectrum of contaminations of the mNGS procedure.

To summarize, I report pioneering data on the performance of mNGS for the diagnosis of meningoencephalitis patients in Vietnam; a resource limited setting. The results shows that in a single assay, mNGS could detect a wide spectrum of neurotropic viruses in CSF samples of meningoencephalitis patients, and thus could potentially replace conventional nucleic acid based diagnostic assays as PCR. The study needed is to determine the clinical implication that real-time metagenomic sequencing may contribute to the diagnosis and management of meningoencephalitis patients, especially in resource-limited settings where pathogen specific assays are limited in number.

Chapter 7: Summary and Future Direction

Despite extensive diagnostic work-up, a significant proportion of patients with CA sepsis or CNS infection have no etiology identified by conventional diagnostic assays such as routine culture, serology and PCR (26–28,173,177,213–215). Additionally, Asia is recognised as a hot spot of emerging infectious diseases, especially those caused by novel viruses, as illustrated by the emergence of Nipah virus in 1997, SARS-CoV-1 in 2002-2003, H5N1 in 2004, and most recently SARS-CoV-2. Moreover, it is estimated that there are more than 320,000 mammalian viruses that have not been characterised (217). Therefore, improving our knowledge about the causes of sepsis and CNS infections and early recognition of emerging pathogens are critical to inform clinical management and outbreak response. Over last decade, mNGS has emerged as a sensitive assay for sequence-independent detection of infectious agents, especially (novel) viral pathogens in clinical samples. Yet, few mNGS associated studies have been conducted in Asia, where the burden of infectious disease is exceptional high. Therefore, my PhD research aimed to:

1. Develop a sensitive viral metagenomic pipeline for sequence-independent detection of a broad range of viral pathogens in clinical samples
2. Explore viral content in patients with sepsis of unknown cause across Southeast Asia
3. Explore viral content in CSF from patients with acute CNS infections of unknown cause sampled from provincial hospitals throughout Vietnam
4. (If relevant), demonstrate proof of causation of recently described viruses/novel virus(es) discovered by metagenomic analysis
5. Explore the utility potential of metagenomics for the diagnosis of patients with central nervous system infections

Herein, I will provide an overview and discussion about my research findings. I also aim to discuss some future directions based on the results I obtained during my PhD research.

1. Findings of the thesis:

In Chapter 2, I set out to develop a sensitive metagenomic assay for detection of a broad range of viral pathogens in clinical samples. After evaluating four possible combinations of sample pretreatment steps, a metagenomic workflow employing a DNase/RNase treatment step of the specimen without sample pre-centrifugation was selected for subsequent metagenomic analysis described in Chapters 3 and 5. Based on data obtained from a reference viral mixture consisting of 20 different viral pathogens (190), my mNGS assay had a comparable sensitivity with that of pipelines developed by my collaborator in the US, Professor Eric Delwart, and a group in Switzerland (190,227). My mNGS assay was also able to detect different viral pathogens including Zika virus, HAV, HBV, HSV DENV and mumps virus in serum or CSF samples. In addition to the detection of viral sequences, the assay mNGS could generate nearly/complete genome sequences of 5/25 and 4/11 viruses in the viral mixture and in the tested clinical samples, respectively. Collectively, within this chapter, I successfully developed a metagenomic pipeline for sensitive detection of a broad range of viral pathogens in clinical samples.

In chapter 3, I applied the mNGS assay developed in Chapter 2 to search for viral agents in samples collected from patients presenting with CA sepsis of unknown origin enrolled in multiple hospitals across Vietnam (n=6) and Thailand (n=4) during 2013-2015. Notably, these patients were examined for a wide range of common causes of CA sepsis using a combination of conventional assays (routine culture, serology, rapid tests, ELISA and PCR) (26), but had no etiology identified. The mNGS detected 26 viral species in 137/466 (29%) and 63/88 (71.6%) samples collected from Vietnam and Thailand, respectively. After viral specific PCR confirmatory testing, the detection rate was reduced to 12.8% (66/466) for Vietnamese samples and 34% (30/88) for samples from Thailand. In total, I successfully identified 22 viral species

known to be infectious to humans in 90 (13.5%) of 665 patients presenting with CA sepsis of unknown cause from both countries. Notably, my analysis demonstrated that the levels of viral richness and diversity were different between Thailand and Vietnam, with more viral species found in Vietnamese patients. In terms of prevalence, EVs (3.6%, 14/386) were more often detected in samples collected from Vietnamese patients, while EBV (10/279, 3.6%) and DENV (3.2%, 9/279) made up the majority of viruses detected in samples collected from patients from Thailand.

Of the viruses detected in patients with CA sepsis, I found sequences of a newly discovered flavivirus species namely HPgV-2 in a serum sample of a patient co-infected with HCV and HIV. Since this represents the first detection of HPgV-2 in Vietnam I conducted further experiment to genetically characterize HPgV-2 and to determine its prevalence, especially among patients with HIV/HCV co-infection. I presented the results of these investigations in Chapter 4. Notably, using real time RT-PCR, I found HPgV-2 RNA in 5/79 HCV/HIV co-infected patients. HPgV-2 RNA however was not detected in other patients' groups, including those with HAV, HBV or HIV mono-infection, and healthy donors. Therefore, my findings support previous work regarding the association between HPgV-2 and HCV/HIV co-infection (269,270,274). To explore the persistence of HPgV-2 in HCV/HIV co-infected people, I used real time PCR to test available longitudinal samples. Subsequently, HPgV-2 RNA was detectable for up to 18 months in 3/5 patients with HCV/HIV co-infection, but was not detected in an available follow-up serum sample collected 14 days after enrollment from a patient with CA infection. These data suggest that HPgV-2 viremia can be transient or persistent for up to 18 months, supporting previous reports (268,269). Whole-genome based phylogenetic analysis revealed that all Vietnamese HPgV-2 isolates were closely related to HPgV-2 strains isolated from the United States and elsewhere, suggesting a global dispersal of HPgV-2. Thus, my

findings have expanded data about the geographic distribution, long-term persistence in patients with HCV/HIV coinfection and genetic diversity of HPgV-2.

The search for viral pathogens in patients with severe infection of unknown origin continued in chapter 5. In this chapter, I used mNGS to define the viral causes in 204 patients presenting with CNS infections of unknown origin. The patients were enrolled from four provincial hospitals across central and southern Vietnam from December 2012 to October 2016. Eight distinct viral species were detected in 107/204 (52.4%) of CSF samples by mNGS. After viral specific PCR confirmatory testing, the detection rate was lowered to 30/204 (14.7%), with enteroviruses being the most common viruses detected (n=23, 11.3%), followed by hepatitis B virus (n=3, 1.5%), human immunodeficiency virus (n=2, 1%), molluscum contagiosum virus and gemycircularvirus (n=1 each, 0.5%). Analysis of available enterovirus sequences generated by mNGS revealed the predominance of echovirus 30 (n=9) followed by enterovirus A71 and enterovirus B80 (1 each). Phylogenetically, the echovirus 30 strains belonged to genogroup V and VIIb.

However, it remains a challenge to identify a viral culprit in patients presenting with community-acquired infections such as sepsis and CNS infections. Indeed, over 80% of the included patients remain undiagnosed after mNGS analysis. Possible contributing factors include the sensitivity of mNGS based diagnosis approach (301–303), the viability of the clinical samples after being stored, albeit at -80^{0}C, for some years (26,277), the presence of the virus in un-analyzed samples (e.g. brain biopsy), and non-viral causes (such as bacteria), which were not the focus of my PhD.

It should however be noted that in several circumstances the mNGS findings described in chapter 3 and 5 cannot directly impute sepsis or CNS causation involving the viruses identified. For example, the detection of blood-borne viruses such as HBV, HIV and HCV in sera samples of CA sepsis patients might represent underlying diseases. Likewise, the detection of these

blood-borne viruses in CSF samples of CNS infections patients might have been a consequence of traumatic tap occurring during the lumbar puncture. Additionally, the detection of CMV, EBV and HHV-6 in respiratory and stool samples may simply reflect the carriage of these viruses in those bodily compartments. Finally, common contaminants such as those listed in Figure 6.3, of which some are likely introduced during the sample collection and/or derived from laboratory reagents, also challenge the interpretation of metagenomic results.

Together with recent reports (175,177,210,212–214) findings from chapters 3 and 4 demonstrate that mNGS could offer a new diagnostic pathway for infectious diseases, especially those caused by diverse pathogens such as CNS infections and sepsis. Therefore, to further inform future research directions, in chapter 6 I explored the utility of mNGS for the diagnosis of CNS infections. To achieve this, I used acute CSF from 66 consecutively treated patients with meningoencephalitis admitted to my hospital, the Hospital for Tropical Diseases in Ho Chi Minh City, Vietnam. For the sensitivity and specificity validation purpose, I included all CSF from the selected patients regardless of routine diagnostic results. To allow for the sensitive detection of a broad range of viral pathogens (DNA and RNA viruses), I used two separate mNGS workflows; one for DNA viruses and one for RNA viruses. Routine diagnosis could identify a virus in 15 (22.7%) patients, including HSV (n=7) and VZV (n=1) by PCR, and mumps virus (n=4), DENV (n=2) and JEV (n=1) by serological diagnosis. mNGS detected HSV, VZV and mumps virus in 5/7, 1/1 and 1/4 of the CSF positive by routine assays, respectively, but DENV and JEV in none of the positive CSF. Additionally, mNGS detected enteroviruses in seven patients of unknown cause. Subsequent expanded PCR testing revealed one additional CSF sample positive mumps virus, which was negative by mNGS assay. Therefore, a combination of routine diagnostic testing, mNGS assays and expanding PCR increased the diagnostic yield from 22.7% (15/66) to 34.8% (23/66). Using PCRs as reference assays, the sensitivity and specificity of mNGS were 74% (14/19) and 66% (31/47), respectively. Collectively, in this final

result chapter, I showed that in a single assay, mNGS could accurately detect a wide spectrum of neurotropic viruses in CSF of meningoencephalitis patients. The study showed the value that real-time sequencing may contribute to the diagnosis and management of meningoencephalitis patients, especially in resource-limited settings where pathogen-specific assays are limited in number.

2. Future direction:

As summarized above, my PhD research has provided significant insights into the epidemiology of CNS infections and CA sepsis in Vietnam and Thailand. It has also advanced our knowledge about the utility potential of mNGS as a single assay for pathogen discovery and infectious disease diagnosis, critical to inform future directions about emerging infections in the region and globally. Heading to the future, my aim is to address some key questions arising as part of my PhD research findings, which I outline below.

2.1. Metagenomics for routine diagnosis and novel pathogen surveillance:

In terms of the utility potential of mNGS for routine diagnosis, the critical question now is to define the extent to which mNGS may influence the management of patients with severe infectious diseases such as CA sepsis, CNS infections and pneumonia. Given the current turnaround time of 2-4 days of mNGS based diagnostic approach, albeit less with MinION based workflow, and the high cost, it is likely that mNGS would be most useful in the scenario where routine diagnosis could not define a causative agent. As such a prospective hospital-based study on patients with CA sepsis, CNS infection or pneumonia who are left undiagnosed after routine diagnostics would be of clinical significance and public health importance to pursue. In doing so, we might be able to detect novel causes of severe infections in hospitalized patients as they emerge. Notably, metagenomic analysis of culture materials and bronchoalveolar lavage fluid from a cluster of hospitalized patients presenting with CA pneumonia of unknown origin in Wuhan, China in late 2019 had led to the discovery of SARS-CoV-2 (47,304). Such prospective

metagenomic studies can now be facilitated by the availability of publically available bioinformatic pipelines (e.g. IDseq (206)), and the in-house mNGS pipeline I have set up during my PhD research.

2.2. Human pegivirus 2:

There are still many unanswered questions about the epidemiology, clinical significance potential, tissue tropism, pathogenesis and evolutionary biology of HPgV-2. Addressing these research questions requires extensive investigation and efforts. I will thus focus on expanding our knowledge about the prevalence and evolutionary biology of HPgV-2. I will base my analysis on patients enrolled in a Wellcome funded clinical trial namely VIETNarms (A strategic post-licensing trial of oral direct-acting antiviral hepatitis C treatment in VIETNam incorporating a novel design with multiple ARMS) (305). The trial sample size is 1000 Vietnamese patients. Using longitudinal samples collected from the study participants, I will focus my analysis on:

Exploring the prevalence and persistence of HPgV-2 in patients with HCV mono-infection and HCV/HIV coinfection:

A combination of RT-PCR and serology will be used to screen for HPgV-2 RNA and antibodies in baseline and longitudinal samples to determine the prevalence of HPgV-2 in patients with HCV mono-infection or HCV/HIV co-infection. The antibody based screening approach could identify individuals with past exposure to HPgV-2 (268,270,274), whereas PCR could help identify those with ongoing viremia.

Sofosbuvir is an effective inhibitor of RNA polymerases from a range of RNA viruses. Therefore, in the absence of an in vitro culture system for HPgV-2, by testing longitudinal samples, there might be an opportunity to assess the impact of direct acting antigens treatment on the duration of HPgV-2 detection in blood of the study participants.

Genetically characterizing HPgV-2:

I will use next-generation based whole-genome deep sequencing approaches available at OUCRU to recover HPgV-2 genomes from positive sera (including baseline and follow-up samples) collected from the study participants. Appropriate phylogenetic approaches (e.g. BEAST) will be utilized to help unravel the evolution and spread of HPgV-2 within Vietnam and beyond. Since deep sequencing enables the detection of thousands of viral haplotypes within any given sample, I will then use this data to re-construct transmission networks where possible, determine the relative rate of evolution of HPgV-2 within individuals and at the population level, and characterize the intra-host evolutionary dynamics of HPgV-2 infection.

3. Summary:

In summary, as part of my PhD research I have successfully set up and implemented an in-house mNGS workflow for the sensitive detection of a broad range of viruses in clinical samples at the clinical laboratory of Oxford University Clinical Research and the Hospital for Tropical Diseases in HCMC, Vietnam. Using this method, I have revealed significant insights into the epidemiology of CA sepsis and CNS infections in Thailand and Vietnam. Yet, it remains a challenge to identify a plausible viral pathogen in patients presenting with these devasting clinical conditions. The results have also contributed expanded data about the epidemiology and genetic diversity of HPgV-2, a recently discovered flavivirus. Additionally, my data demonstrated that mNGS can be as a pan-viral assay for infectious disease diagnosis and novel virus surveillance. Therefore, heading to the future it is critical to conduct further research to assess the extent to which real-time mNGS could influence the management of patients with severe infections such as pneumonia, sepsis and CNS infections. This would in turn allow for early detection of novel viral pathogens as they emerge, critical to inform outbreak response. As for HPgV-2, given the global burden of HCV and HIV infections, it is of equal importance to gain further insights into the prevalence and, evolutionary biology of this novel virus.

References

1. Angus DC. The Lingering Consequences of Sepsis. Jama. 2010;304(16):1833.
2. World Health Organization. Global report on the epidemiology and burden of sepsis: current evidence, identifying gaps and future directions. Who. 2020. 56 p.
3. Bone RC, Balk RA, Cerra FB, Dellinger RP, Fein AM, Knaus WA, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. Chest. 1992;101(6):1644–55.
4. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. Intensive Care Med. 2003;29(4):530–8.
5. Dellinger RP, Levy MM, Rhodes A, Bs MB, Annane D, Gerlach H, et al. Surviving Sepsis Campaign : International Guidelines for Management of Severe Sepsis and Septic Shock : 2012. Crit Care Med. 2013;41(2):580–637.
6. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock. Jama. 2016;315(8):801–10.
7. Dondorp AM, Dünser MW, Schultz MJ. Sepsis management in resource-limited settings. Sepsis Manag Resour Settings. 2019;1–216.
8. Reinhart K, Daniels R, Kissoon N, Machado FR, Schachter RD, Finfer S. Recognizing Sepsis as a Global Health Priority — A WHO Resolution. N Engl J Med. 2017;377(5):414–7.
9. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. Lancet. 2020;395(10219):200–11.
10. Vincent JL, Marshall JC, Ñamendys-Silva SA, François B, Martin-Loeches I, Lipman J, et al. Assessment of the worldwide burden of critical illness: The Intensive Care Over Nations (ICON) audit. Lancet Respir Med. 2014;2(5):380–6.
11. Danai P, Martin GS. Epidemiology of sepsis: Recent advances. Curr Infect Dis Rep. 2005;7(5):329–34.
12. Fleischmann C, Scherag A, Adhikari NK, Hartog CS. Assessment of global incidence and mortality of hospital-treated sepsis.Current estimates and limitations. Am J Respir Crit Care Med. 2016;193(3):259–72.
13. Hall MJ, Williams SN, DeFrances CJ, Golosinskiy A. Inpatient care for septicemia or sepsis: a challenge for patients and hospitals. NCHS Data Brief. 2011;(62):1–8.
14. Lagu T, Rothberg MB, Shieh MS, Pekow PS, Steingrub JS, Lindenauer PK. Hospitalizations, costs, and outcomes of severe sepsis in the United States 2003 to 2007. Crit Care Med. 2012;40(3):754–61.
15. Fleischmann C, Thomas-Rueddel DO, Hartmann M, Hartog CS, Welte T, Heublein S, et al. Hospital Incidence and Mortality Rates of sepsis. Dtsch Arztebl Int. 2016;113(10):159–66.
16. Bouza C, López-Cuadrado T, Saz-Parkinson Z, María J, -Blanco A. Epidemiology and recent trends of severe sepsis in Spain: a nationwide population-based analysis (2006-2011). 2015.
17. Harrison DA, Welch CA, Eddleston JM. The epidemiology of severe sepsis in England, Wales and Northern Ireland, 1996 to 2004: secondary analysis of a high quality clinical database, the ICNARC Case Mix Programme Database. Crit Care Med. 2006;10.
18. Kaukonen KM, Bailey M, Suzuki S, Pilcher D, Bellomo R. Mortality related to severe

sepsis and septic shock among critically ill patients in Australia and New Zealand, 2000-2012. JAMA - J Am Med Assoc. 2014;311(13):1308–16.
19. Lewis JM, Abouyannis M, Katha G, Nyirenda M, Chatsika G, Feasey NA, et al. Population incidence and mortality of sepsis in an urban African setting 2013-2016. Clin Infect Dis. 2019;71:2013–6.
20. Hantrakun V, Somayaji R, Teparrukkul P, Boonsri C. Clinical epidemiology and outcomes of community acquired infection and sepsis among hospitalized patients in a resource limited setting in Northeast Thailand : A prospective observational study (Ubon-sepsis). PLoS One. 2018;1–14.
21. Khairul A, Purba R, Mariana N, Aliska G, Hadi S. The burden and costs of sepsis and reimbursement of its treatment in a developing country : An observational study on focal infections in Indonesia. Int J Infect Dis. 2020;96:211–8.
22. Martin GS. Sepsis, severe sepsis and septic shock: changes in incidence, pathogens and outcomes. Expert Rev Anti Infect Ther. 2013;10(6):701–6.
23. Arefian H, Heublein S, Scherag A, Brunkhorst FM, Younis MZ, Moerer O, et al. Hospital-related cost of sepsis: A systematic review. J Infect. 2017;74(2):107–17.
24. Torio CM, Andrews RM. National Inpatient Hospital Costs: the Most Expensive Conditions by Payer, 2011.Statistical Brief #160. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Heal Cost Util Proj. 2013;31(1):1–12.
25. Mannino M, Martin GS. The role of infection and comorbidity: Factors that influence disparities in sepsis. Crit Care Med. 2014;34(10):2576–82.
26. Southest Asia Infectious Disease Clinical Research. Causes and outcomes of sepsis in southeast Asia: a multinational multicentre cross-sectional study. Lancet Glob Heal. 2017;5(2):e157–67.
27. Prost N De, Razazi K, Brun-buisson C. Unrevealing culture-negative severe sepsis. Crit Care. 2013;17:1001.
28. Phua J, Ngerng WJ, See KC, Tay CK, Kiong T, Lim HF, et al. Characteristics and outcomes of culture-negative versus culture-positive severe sepsis. Crit Care. 2013;17:R002.
29. Natanson C, Suflfredini AF, Danner RL, Cunnion RE, Ognibene FP. Septic Shock in Humans Advances in the Understanding of Pathogenesis , Cardiovascular. Ann Intern Med. 1990;113(3):227–42.
30. Glauser MP. Antibiotics in sepsis. Intensive Care Med. 2001;27:33–48.
31. Moss M. The Epidemiology of Sepsis in the United States from 1979 through 2000. 2003;1546–54.
32. Opal SM, Garber GE, Larosa SP, Maki DG, Freebairn RC, Kinasewitz GT, et al. Systemic Host Responses in Severe Sepsis Analyzed by Causative Microorganism and Treatment Effects of Drotrecogin Alfa (Activated). Clin Infect Dis. 2003;37.
33. Zaidi AKM, Thaver D, Ali SA. Pathogens Associated With Sepsis in Newborns and Young Infants in Developing Countries. 2009;28(1):10–8.
34. Topeli A, Ünal S, Akalin HE. Risk factors influencing clinical outcome in Staphylococcus aureus bacteraemia in a Turkish University Hospital. Int J Antimicrob Agents. 2000;14(1):57–63.
35. Turnidge J. Impact of antibiotic resistance on the treatment of sepsis. Scand J Infect Dis. 2003;35(9):677–82.
36. Delaloye J, Calandra T. Invasive candidiasis as a cause of sepsis in the critically ill patient. Virulence. 2014;5(1):161–9.
37. Xie G, Fang X, Fang Q, Wu X, Jin Y, Wang J, et al. Impact of invasive fungal infection on outcomes of severe sepsis : a multicenter matched cohort study in critically ill surgical

patients. Crit Care. 2008;12:1–9.
38. Ranieri VM, Reinhart K, Gerlach H, Moreno R. Sepsis in European intensive care units: Results of the SOAP study*. Crit Care Med. 2006;34(2):344–53.
39. Haouache H, Goldgran-toledano D, Allaouchiche B, Azoulay E. Outcomes in severe sepsis and patients with septic shock: Pathogen species and infection sites are not associated with mortality*. Crit Care Med. 2011;39(8):1886–95.
40. Wilder-smith A, Ooi E. Update on Dengue : Epidemiology , Virus Evolution , Antiviral Drugs , and Vaccine Development. Curr Infect Dis Rep. 2010;12:157–64.
41. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013;496(7446):504–7.
42. Teparrukkul P, Hantrakun V, Day NPJ, West TE, Limmathurotsakul D. Management and outcomes of severe dengue patients presenting with sepsis in a tropical country. PLoS One. 2017;12(4):1–13.
43. Shane AL, Sánchez PJ, Stoll BJ. Seminar Neonatal sepsis. Lancet. 2017;6736(17):1–11.
44. Wolthers KC, Benschop KSM, Schinkel J, Molenkamp R, Bergevoet RM, Spijkerman IJB, et al. Human Parechoviruses as an Important Viral Cause of Sepsislike Illness and Meningitis in Young Children. Clin Infect Dis. 2008;47:358–63.
45. Cillóniz C, Dominedò C, Magdaleno D, Ferrer M. Pure Viral Sepsis Secondary to Community-Acquired Pneumonia in Adults : Risk and Prognostic Factors. J Infect Dis. 2019;1–6.
46. Ortiz JR, Neuzil KM, Shay DK, Rue TC, Neradilek MB, Zhou H, et al. The Burden of Influenza-Associated Critical Illness Hospitalizations*. Crit Care Med. 2014;42(11):2325–32.
47. Acuti C, Elena M, Cappadona R, Bravi F, Mantovani L, Manzoli L. SARS-CoV-2 pandemic : An overview. Adv Biol Regul. 2020;77:1–11.
48. Eifan SA, Nour I, Hanif A, Zamzam AMM, Mohammed S. A pandemic risk assessment of middle east respiratory syndrome coronavirus (MERS-CoV) in Saudi Arabia. Saudi J Biol Sci. 2017;24(7):1631–8.
49. Cherry JD, Krogstad P. SARS : The First Pandemic of the 21 st Century. Pediatr Res. 2004;56(1):1–5.
50. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan , China : a retrospective cohort study. Lancet. 2020;395:1054–62.
51. WHO. Neurological disorders public health challenges. Medicine (Baltimore). 2006;229.
52. Riddell J, Shuman EK. Epidemiology of Central Nervous System Infection. Neuroimaging Clin N Am. 2012;22(4):543–56.
53. Levin SN, Lyons JL. Infections of the Nervous System. Am J Med. 2018;131(1):25–32.
54. Kim KS. Acute bacterial meningitis in infants and children. Lancet Infect Dis. 2010;10(1):32–42.
55. van de Beek D, de Gans J, Spanjaard L, Weisfelt M, Reitsma JB, Vermeulen M. Clinical Features and Prognostic Factors in Adults with Bacterial Meningitis. N Engl J Med. 2004;351(18):1849–59.
56. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2095–128.
57. McGill F, Griffiths MJ, Solomon T. Viral meningitis: Current issues in diagnosis and treatment. Curr Opin Infect Dis. 2017;30(2):248–56.
58. Rice P. Viral meningitis and encephalitis. Am Acad Neurol. 2018;24(5):1284–97.

59. Wright WF, Pinto CN, Palisoc K, Baghli S. Viral (aseptic) meningitis: A review. J Neurol Sci. 2019;398:176–83.
60. Ihekwaba UK, Kudesia G, McKendrick MW. Clinical features of viral meningitis in adults: Significant differences in cerebrospinal fluid findings among herpes simplex virus, varicella zoster virus, and enterovirus infections. Clin Infect Dis. 2008;47(6):783–9.
61. Omland LH, Vestergaard BF, Wandall JH. Herpes simplex virus type 2 infections of the central nervous system: A retrospective study of 49 patients. Scand J Infect Dis. 2008;40(1):59–62.
62. Michos AG, Syriopoulou VP, Hadjichristodoulou C, Daikos GL, Lagona E, Douridas P, et al. Aseptic menimgitis in children: Analysis of 506 cases. PLoS One. 2007;2(8):e674.
63. Molyneux E, Riordan FAI, Walsh A. Acute bacterial meningitis in children presenting to the Royal Liverpool Children's Hospital, Liverpool, UK and the Queen Elizabeth Central Hospital in Blantyre, Malawi: A world of difference. Ann Trop Paediatr. 2006;26:29–37.
64. Kumar R. Understanding and managing acute encephalitis. F1000Research. 2020;9:60.
65. Granerod J, Tam CC, Crowcroft NS, Davies NWS, Borchert M, Thomas SL. Challenge of the unknown: A systematic review of acute encephalitis in non-outbreak situations. Neurology. 2010;75(10):924–32.
66. Somand D, Meurer W. Central Nervous System Infections. Emerg Med Clin North Am. 2009;27(1):89–100.
67. Davies NWS, Sharief MK, Howard RS. Infection-associated encephalopathies-their investigation, diagnosis, and treatment. J Neurol. 2006;253(7):833–45.
68. Venkatesan A, Tunkel AR, Bloch KC, Lauring AS, Sejvar J, Bitnun A, et al. Case definitions, diagnostic algorithms, and priorities in encephalitis: Consensus statement of the international encephalitis consortium. Clin Infect Dis. 2013;57(8):1114–28.
69. Lancaster E, Dalmau J. Neuronal autoantigens—pathogenesis, associated disorders and antibody testing. Nat Rev Neurol. 2012;8(7):380–90.
70. Granerod J, Ambrose HE, Davies NWS, Clewley JP, Walsh AL, Morgan D, et al. Causes of encephalitis and differences in their clinical presentations in England : a multicentre , population-based prospective study. Lancet Infect Dis. 2010;10(12):835–44.
71. Dalmau J, Tu E, Rossi JE, Voloschin A, Baehring JM, Shimazaki H, et al. Paraneoplastic Anti – N -methyl- D -aspartate Receptor Encephalitis Associated with Ovarian Teratoma. Ann Neurol. 2007;61:25–36.
72. Abrantes F, Toso FF, Povoas OG, Hoftberger R. Autoimmune encephalitis : a review of diagnosis and treatment. Arq Neuropsiquiatr. 2018;76(1):41–9.
73. Graus F, Titulaer M., Balu R, Benseler S, Bien C., Cellucci T, et al. A clinical approach to diagnosis of autoimmune encephalitis. Lancet Neurol. 2016;15(4):391–404.
74. Feigin VL, Krishnamurthi R V., Theadom AM, Abajobir AA, Mishra SR, Ahmed MB, et al. Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Neurol. 2017;16(11):877–97.
75. Zunt JR, Kassebaum NJ, Blake N, Glennie L, Wright C, Nichols E, et al. Global, regional, and national burden of meningitis, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2018;17(12):1061–82.
76. Robertson FC, Lepard JR, Mekary RA, Davis MC, Yunusa I, Gormley WB, et al. Epidemiology of central nervous system infectious diseases: A meta-analysis and systematic review with implications for neurosurgeons worldwide. J Neurosurg. 2018;130(4):1107–26.
77. Ford H, Wright J. Bacterial meningitis in Swaziland: An 18 month prospective study of

its impact. J Epidemiol Community Health. 1994;48(3):276–80.
78. Chan YC, Ong BKC, Kumarasinghe G. Meningitis in a Singaporean Teaching Hospital . A Seven-Year Overview (1993-2000). Singapore Med J. 2002;43(12):632–6.
79. Scarborough M, Thwaites GE. The diagnosis and management of acute bacterial meningitis in resource-poor settings. Lancet Neurol. 2008;7(7):637–48.
80. McIntyre PB, O'Brien KL, Greenwood B, Van De Beek D. Effect of vaccines on bacterial meningitis worldwide. Lancet. 2012;380(9854):1703–11.
81. Basri R, Zueter AR, Mohamed Z, Alam MK, Norsa'Adah B, Hasan SA, et al. Burden of bacterial meningitis: A retrospective review on laboratory parameters and factors associated with death in Meningitis, Kelantan Malaysia. Nagoya J Med Sci. 2015;77(1–2):59–68.
82. Shaker R, Fayad D, Dbaibo G. Challenges and opportunities for meningococcal vaccination in the developing world. Hum Vaccines Immunother. 2018;14(5):1084–97.
83. Kakooza-Mwesige A, Tshala-Katumbay D, Juliano SL. Viral infections of the central nervous system in Africa. Brain Res Bull. 2019;145(May 2018):2–17.
84. Kupila L, Vuorinen T, Vainionpää R, Hukkanen V, Marttila RJ, Kotilainen P. Etiology of aseptic meningitis and encephalitis in an adult population. Neurology. 2006;66(1):75–80.
85. Rantakallio P, Leskinen M, von Wendt L. Incidence and prognosis of central nervous system infections in a birth cohort of 12 000 children. Scand J Infect Dis. 1986;18(4):287–94.
86. Khetsuriani N, Quiroz ES, Holman RC, Anderson LJ. Viral meningitis-associated hospitalizations in the United States, 1988-1999. Neuroepidemiology. 2003;22(6):345–52.
87. Soares CN, Cabral-Castro MJ, Peralta JM, De Freitas MRG, Zalis M, Puccioni-Sohler M. Review of the etiologies of viral meningitis and encephalitis in a dengue endemic region. J Neurol Sci. 2011;303(1–2):75–9.
88. Meyer HM, Johnson RT, Crawford IP, Dascomb HE, Rogers NG. Central nervous system syndromes of "viral" etiology. A study of 713 cases. Am J Med. 1960;29(2):334–47.
89. Thompson C, Kneen R, Riordan A, Kelly D, Pollard AJ. Encephalitis in children. Arch Dis Child. 2012;97(2):150–61.
90. Greenlee JE. Encephalitis and postinfectious encephalitis. Contin Lifelong Learn Neurol. 2012;18(6):1271–89.
91. Tan L Van, Thai LH, Phu NH, Nghia HDT, Chuong L Van, Sinh DX, et al. Viral Aetiology of Central Nervous System Infections in Adults Admitted to a Tertiary Referral Hospital in Southern Vietnam over 12 Years. PLoS Negl Trop Dis. 2014;8(8):e3127.
92. Solomon T, Michael BD, Smith PE, Sanderson F, Davies NWS, Hart IJ, et al. Management of suspected viral encephalitis in adults - Association of British Neurologists and British Infection Association National Guidelines. J Infect. 2012;64(4):347–73.
93. Vora NM, Holman RC, Mehal JM, Steiner CA, Blanton J, Sejvar J. Burden of encephalitis-associated hospitalizations in the United States, 1998-2010. Neurology. 2014;82(5):443–51.
94. Ho NT, Hoang VMT, Le NNT, Nguyen DT, Tran A, Kaki D, et al. A spatial and temporal analysis of paediatric central nervous system infections from 2005 to 2015 in Ho Chi Minh City, Vietnam. Epidemiol Infect. 2017;145(15):3307–17.
95. Ho Dang Trung N, Le Thi Phuong T, Wolbers M, Nguyen van Minh H, Nguyen Thanh V, Van MP, et al. Aetiologies of central nervous system infection in Viet Nam: A prospective provincial hospital-based descriptive surveillance study. PLoS One.

2012;7(5):e37875.
96. Taylor WR, Nguyen K, Nguyen D, Nguyen H, Horby P, Nguyen HL, et al. The Spectrum of Central Nervous System Infections in an Adult Referral Hospital in Hanoi, Vietnam. PLoS One. 2012 Aug 30;7(8).
97. Glaser CA, Honarmand S, Anderson LJ, Schnurr DP, Forghani B, Cossen CK, et al. Beyond Viruses : Clinical Profiles and Etiologies Associated with Encephalitis. Clin Infect Dis. 2006;43(12):1565–77.
98. Olsen SJ, Campbell AP, Supawat K, Liamsuwan S, Chotpitayasunondh T, Laptikulthum S, et al. Infectious Causes of Encephalitis and Meningoencephalitis in Thailand, 2003–2005. Emerg Infect Dis. 2015;21(2):280–9.
99. van Tan L, Qui PT, Ha DQ, Hue NB, Bao LQ, van Cam B, et al. Viral etiology of encephalitis in children in Southern Vietnam: Results of a one-year prospective descriptive study. PLoS Negl Trop Dis. 2010;4(10).
100. Taylor WR, Nguyen K, Nguyen D, Nguyen H, Horby P, Nguyen HL, et al. The Spectrum of Central Nervous System Infections in an Adult Referral Hospital in Hanoi, Vietnam. PLoS One. 2012;7(8):1–8.
101. van Tan L, Qui PT, Ha DQ, Hue NB, Bao LQ, van Cam B, et al. Viral etiology of encephalitis in children in Southern Vietnam: Results of a one-year prospective descriptive study. PLoS Negl Trop Dis. 2010;4(10):e854.
102. Bookstaver PB, Mohorn PL, Shah A, Tesh LD, Quidley AM, Kothari R, et al. Management of Viral Central Nervous System Infections: A Primer for Clinicians. J Cent Nerv Syst Dis. 2017;9:1–12.
103. Ahlbrecht AJ, Hillebrand LK, Ganzenmueller T, Heim A, Stangel M, Kurt-wolfram S. Cerebrospinal fluid features in adults with enteroviral nervous system infection. Int J Infect Dis. 2018;
104. Vollbach S, Müller A, Drexler JF, Simon A, Drosten C, Eis-hübinger AM. Prevalence , type and concentration of human enterovirus and parechovirus in cerebrospinal fluid samples of pediatric patients over a 10-year period : a retrospective study. Virol J. 2015;1–6.
105. Holmes CW, Koo SSF, Osman H, Wilson S, Xerry J, Gallimore CI, et al. Predominance of enterovirus B and echovirus 30 as cause of viral meningitis in a UK population. J Clin Virol. 2016;81:90–3.
106. B'Krong NTTC, Minh NNQ, Qui PT, Chau TTH, Nghia HDT, Do LAH, et al. Enterovirus serotypes in patients with central nervous system and respiratory infections in Viet Nam 1997–2010. Virol J. 2018;15:69.
107. Puenpa J, Wanlapakorn N, Vongpunsawad S, Poovorawan Y. The History of Enterovirus A71 Outbreaks and Molecular Epidemiology in the Asia-Pacific Region. J Biomed Sci. 2019;26:75.
108. Mcminn PC. Enterovirus 71 in the Asia-Pacific region : An emerging cause of acute neurological disease in young children. Neurol J Southeast Asia. 2003;8:57–63.
109. Messacar K, Asturias EJ, Hixon AM, Van Leer-Buter C, Niesters HGM, Tyler KL, et al. Enterovirus D68 and acute flaccid myelitis—evaluating the evidence for causality. Lancet Infect Dis. 2018;18(8):e239–47.
110. Logan SAE, MacMahon E. Viral meningitis. Bmj. 2008;336(7634):36–40.
111. Honda H, Warren DK. Central Nervous System Infections: Meningitis and Brain Abscess. Infect Dis Clin North Am. 2009;23(3):609–23.
112. Rotbart HA, Brennan PJ, Fife KH, Romero JR, Johanna A, Mckinlay MA, et al. Enterovirus Meningitis in Adults. Clin Infect Dis. 1998;27(4):896–8.
113. Meningitis E, Meningitis MS. Herpes Simplex Virus Infections of the Central Nervous

System : Encephalitis and Meningitis , Including Mollaret ' s. Herpes. 2004;11(Suppl 2):57A–64A.
114. Kallio-laine K, Seppänen M, Lappalainen M, Valtonen V. Recurrent Lymphocytic Meningitis Positive for Herpes Simplex Virus Type 2. Emerg Infect Dis. 2009;15(7):1119–22.
115. Ben Abid F, Abukhattab M, Ghazouani H, Khalil O, Gohar A, Al Soub H, et al. Epidemiology and clinical outcomes of viral central nervous system infections. Int J Infect Dis. 2018;73:85–90.
116. Hoffman GM, Ghanayem NS, Amlie-lefond CM, Schwabe MJ, Chusid MJ, Rupprecht CE, et al. Survival after Treatment of Rabies with Induction of Coma. N Engl J Med. 2005;352(24):2508–14.
117. Hampson K, Coudeville L, Lembo T, Sambo M, Kieffer A, Attlan M, et al. Estimating the Global Burden of Endemic Canine Rabies. PLoS Negl Trop Dis. 2015;9(4):1–20.
118. Lumlertdacha B, Guerra M, Meltzer MI, Dhankhar P. Human Rabies Prevention — United States , 2008 Recommendations of the Advisory Committee on Immunization Practices. MMWR Recomm Rep. 2008;57(RR-3):1–28.
119. Pierson TC, Diamond MS. The continued threat of emerging flaviviruses. Nat Microbiol. 2020;5:796–812.
120. WHO. Dengue and severe dengue. World Health Organization. 2020. p. 1–13.
121. Campbell GL, Hills SL, Fischer M, Jacobson JA, Hoke CH, Hombach JM, et al. Estimated global incidence of Japanese encephalitis: A systematic review. Bull World Health Organ. 2011;89(10):766–74.
122. Mostashari F, Bunning ML, Kitsutani PT, Singer DA, Nash D, Cooper MJ, et al. Epidemic West Nile encephalitis , New York , 1999 : results of a household-based seroepidemiological survey. Lancet. 2001;358:261–4.
123. Leary DRO, Marfin AA, Montgomery SP, Kipp AM, Lehman JA, Biggerstaff BJ, et al. The Epidemic of West Nile Virus in the United States, 2002. VECTOR-BORNE ZOONOTIC Dis. 2004;4(1):61–70.
124. CDC. West Nile Virus Activity — United States , 2009. 2010;59(25).
125. Doughty CT, Yawetz S, Lyons J. Emerging Causes of Arbovirus Encephalitis in North America : Powassan , Chikungunya , and Zika Viruses. Curr Neurol Neurosci Rep. 2017;17:12.
126. Vodušek VF, Vizjak A, Ph D, Pižem J, Ph D. Zika Virus Associated with Microcephaly. N Engl J Med. 2016;374:951–8.
127. Victor A, Coelho C. Microcephaly Prevalence in Infants Born to Zika Virus-Infected Women : A Systematic Review and Meta-Analysis. Int J Mol Ciences. 2017;18:1714.
128. Clemmons NS, Gastanaduy PA, Fiebelkorn AP, Redd SB, Wallace GS. Measles — United States, January 4–April 2, 2015. MMWRMorbMortalWkly Rep. 2015;64(14):373–376.
129. Zeng SZ, Zhang B ZY. Identification of 12 cases of acute measles encephalitis without rash. Clin Infect Dis. 2016;63(12):1630–3.
130. Campbell H, Andrews N, Brown KE, Miller E. Review of the effect of measles vaccination on the epidemiology of SSPE. Int J Epidemiol. 2007;36(6):1334–48.
131. Paulson GW. Subacute sclerosing panencephalitis. Pediatr Heal Med Ther. 2018;9:67–71.
132. Ibrahim SH, Amjad N, Saleem AF, Chand P, Rafique A, Humayun KN. The upsurge of SSPE-A reflection of national measles immunization status in Pakistan. J Trop Pediatr. 2014;60(6):449–53.
133. States MU, May J, Gastañaduy PA, Redd SB, Fiebelkorn AP, Rota JS, et al. Measles —

United States, January 1–May 23, 2014. MMWR Morb Mortal Wkly Rep. 2014;63(22):496–9.
134. Choisy M, Tung TS, Thi N, Diep N, Hien NT, Le MQ, et al. Sero-Prevalence Surveillance to Predict Vaccine- Preventable Disease Outbreaks ; A Lesson from the 2014 Measles Epidemic in Northern Vietnam. Open Forum Infect Dis. 2014;6(3):ofz030.
135. Tyor W, Harrison T. Mumps and rubella. 1st ed. Vol. 123, Neurovirology. Elsevier B.V.; 2014. 591–600 p.
136. CDC. Mumps Outbreak — New York and New Jersey , June 2009 – January 2010. MMWR Morb Mortal Wkly Rep. 2010;59(5):125–9.
137. Chua KB. Introduction : Nipah Virus — Discovery and Origin. Curr Top Microbiol Immunol. 2012;359:1–9.
138. Murray K, Selleck P, Hooper P, Hyatt A, Gould A, Gleeson L, et al. A Morbillivirus That Caused Fatal Disease in Horses and Humans. Science (80-). 1995;268:94–7.
139. Wildlife Health Australia. Hendra virus and Australian wildlife. WHA Fact Sheet: Hendra virus in flying-foxes in Australia. 2017.
140. Fenollar F, Raoult D. Molecular diagnosis of bloodstream infections caused by non-cultivable bacteria. Iternational J Antimicrob Agents. 2007;30S:S7–15.
141. Sinha M, Jupe J, Mack H, Coleman TP, Lawrence SM, Fraley I. Emerging Technologies for Molecular Diagnosis of Sepsis. J Clin Microbiol. 2018;31(2):e00089-17.
142. Fan S, Miller NS, Lee J, Remick DG. Diagnosing Sepsis – The Role of Laboratory Medicine. Clin Chim Acta. 2016;460:203–10.
143. Bs RAG, Ascp MT, Spitzer ED, Rn JB, Bsn CB, Bsn RD, et al. Multidisciplinary team review of best practices for collection and handling of blood cultures to determine effective interventions for increasing the yield of true-positive bacteremias , reducing contamination , and eliminating false-positive central line. Am J Infect Control. 2015;43(11):1222–37.
144. Ilstrup DM, Ii JAW. The Importance of Volume of Blood Cultured in the Detection of Bacteremia and Fungemia. Diagn Microbiol Infect Dis. 1983;1:107–10.
145. Martinez RM, Wolk DM. Bloodstream Infections. Microbiol Spectr. 2016;4(4):1–34.
146. Thomson RB, Bertram H. Laboratory diagnosis of central nervous system infections. Infect Dis Clin North Am. 2001;15(4):1047–71.
147. Steiner I, Budka H, Chaudhuri A, Koskiniemi M, Sainio K, Salonen O, et al. Viral meningoencephalitis : a review of diagnostic methods and guidelines for management. Eur J Neurol. 2010;17:999–1009.
148. He T, Kaplan S, Kamboj M, Tang Y. Laboratory Diagnosis of Central Nervous System Infection. Curr Infect Dis Rep. 2016;18:35.
149. Ginocchio CC, Zhang F, Malhotra A, Manji R, Sillekens P, Foolen H, et al. Development , Technical Performance , and Clinical Evaluation of a NucliSens Basic Kit Application for Detection of Enterovirus RNA in Cerebrospinal Fluid. J Clin Microbilogy. 2005;43(6):2616–23.
150. Polage CR, Petti CA. Assessment of the Utility of Viral Culture of Cerebrospinal Fluid. Clin Infect Dis. 2006;43:1578–9.
151. Solomon T. Flavivirus Encephalitis. N Engl J Med. 2004;351:370–8.
152. Chabierski S, Barzon L, Papa A, Niedrig M, Bramson JL, Richner JM, et al. Distinguishing West Nile virus infection using a recombinant envelope protein with mutations in the conserved fusion-loop. BMC Infect Dis. 2014;14:246.
153. Peters RPH, Agtmael MA Van, Danner SA, Savelkoul PHM, Christina MJE. Review New developments in the diagnosis of bloodstream infections. Lancet Infect Dis. 2004;4(12):751–60.

154. Gutierrez M, Emmanuel PJ. Expanding Molecular Diagnostics for Central Nervous System Infections. Adv Pediatr. 2018;
155. Shaker OG, Abdelhamid N. Detection of enteroviruses in pediatric patients with aseptic meningitis. Clin Neurol Neurosurg. 2015;129:67–71.
156. Rowley AH, Whitley RJ, Lakeman FD, Wolinsky SM. Rapid detection of herpes-simplex-virus DNA in cerebrospinal fluid of patients with herpes simplex encephalitis. Lancet. 1990;335:440–1.
157. Mchugh MP, Gray SJ, Kaczmarski EB, Guiver M. Reduced turnaround time and improved diagnosis of invasive serogroup B Neisseria meningitidis and Streptococcus pneumoniae infections using a lyophilized quadruplex quantitative PCR. J Med Microbiol. 2015;64:1321–8.
158. Deng J, Pei J, Gou H, Ye Z, Liu C, Chen J. Rapid and simple detection of Japanese encephalitis virus by reverse transcription loop-mediated isothermal amplification combined with a lateral flow dipstick. J Virol Methods. 2015;213:98–105.
159. Lakeman FD, Whitley RJ. Diagnosis of Herpes Simplex Encephalitis : Application of Polymerase Chain Reaction to Cerebrospinal Fluid from Brain-Biopsied Patients and Correlation with Disease. J Infect Dis. 1995;171:857–63.
160. Glass N, Nelson HD, Huffman L. Screening for Genital Herpes Simplex : Brief Update for the U . S . Preventive Services Task Force Types of Genital HSV. Am Fam Physician. 2005;72(8):1557–61.
161. Mardis ER. DNA sequencing technologies : 2006 – 2016. Nat Protoc. 2017;12(2):213–8.
162. Gupta AK, Gupta UD. Next Generation Sequencing and Its Applications. Animal Biotechnology: Models in Discovery and Translation. Elsevier; 2014. 345–367 p.
163. Kulski JK. Next-Generation Sequencing — An Overview of the History , Tools , and " Omic " Applications. In: INTECH. 2015. p. 3–60.
164. Illumina. Explore Illumina sequencing technology Massively parallel sequencing with optimized SBS chemistry. Illumina. p. https://sapac.illumina.com/science/technology/next.
165. Jerome H, Taylor C, Sreenu VB, Klymenko T, Da A, Filipe S, et al. Metagenomic next-generation sequencing aids the diagnosis of viral infections in febrile returning travellers. J Infect. 2019;79:383–8.
166. Tian J, Pei Y, Yuan M, Zhang Y, Dai F, Liu Y, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579:265–269.
167. Brown JR, Morfopoulou S, Hubb J, Emmett WA, Ip W, Shah D, et al. Astrovirus VA1/HMO-C: An increasingly recognized neurotropic pathogen in immunocompromised patients. Clin Infect Dis. 2015;60(6):881–8.
168. Stremlau MH, Andersen KG, Folarin OA, Grove JN, Odia I, Ehiane PE, et al. Discovery of Novel Rhabdoviruses in the Blood of Healthy Individuals from West Africa. PLoS Negl Trop Dis. 2015 Mar;9(3):e0003631.
169. Phan TG, Mendoza J del V, Sadeghi M, Altan E, Deng X, Delwart E. Sera of Peruvians with fever of unknown origins include viral nucleic acids from non-vertebrate hosts. Virus Genes. 2018 Feb;54(1):33.
170. Loh J, Zhao G, Presti RM, Holtz LR, Finkbeiner SR, Droit L, et al. Detection of Novel Sequences Related to African Swine Fever Virus in Human Serum and Sewage. J Virol. 2009 Dec 15;83(24):13019–25.
171. Takeuchi S, Kawada J, Horiba K, Okuno Y, Okumura T, Suzuki T, et al. Metagenomic analysis using next-generation sequencing of pathogens in bronchoalveolar lavage fluid from pediatric patients with respiratory failure. Sci Rep. 2019 Dec 9;9(1):12909.
172. Perlejewski K, Popiel M, Laskus T, Nakamura S, Motooka D, Stokowy T, et al. Next-

generation sequencing (NGS) in the identification of encephalitis-causing viruses: Unexpected detection of human herpesvirus 1 while searching for RNA pathogens. J Virol Methods. 2015 Dec 15;226:1–6.
173. Kufner V, Plate A, Schmutz S, Braun DL, Günthard HF, Capaul R, et al. Two Years of Viral Metagenomics in a Tertiary Diagnostics Unit: Evaluation of the First 105 Cases. Genes (Basel). 2019;10(9).
174. Linsuwanon P, Poovorawan Y, Li L, Deng X, Vongpunsawad S, Delwart E. The Fecal Virome of Children with Hand, Foot, and Mouth Disease that Tested PCR Negative for Pathogenic Enteroviruses. PLoS One. 2015;10(8):e0135573.
175. Brown JR, Bharucha T, Breuer J. Encephalitis diagnosis using metagenomics: application of next generation sequencing for undiagnosed cases. J Infect. 2018;76(3):225–40.
176. Ajogbasile F V., Oguzie JU, Oluniyi PE, Eromon P, Uwanibe J, Mehta SB, et al. Real-time Metagenomic Analysis of Undiagnosed Fever Cases Unveils a Yellow Fever Outbreak in Edo State, Nigeria. Sci Rep. 2020;10(1):3180.
177. Tschumi F, Schmutz S, Kufner V, Heider M, Pigny F, Schreiner B, et al. Meningitis and epididymitis caused by Toscana virus infection imported to Switzerland diagnosed by metagenomic sequencing: a case report. BMC Infect Dis. 2019 Dec 8;19(1):591.
178. Golan D, Medvedev P. Using state machines to model the Ion Torrent sequencing process and to improve read error rates. Bioinformatics. 2013;29:344–51.
179. Marascio N, Pavia G, Strazzulla A, Dierckx T, Cuypers L, Vrancken B, et al. Detection of Natural Resistance-Associated Substitutions by Ion Semiconductor Technology in HCV1b Positive , Direct-Acting Antiviral Agents-Naïve Patients. Int J Mol Sci. 2016;17:1416.
180. Nougairede A, Bichaud L, Thiberville S, Ninove L, Zandotti C, Lamballerie X De, et al. Isolation of Toscana Virus from the Cerebrospinal Fluid. VECTOR-BORNE ZOONOTIC Dis. 2013;13(9):1316.
181. Csabai Z, Sharon D, Snyder M, Biology PD. Characterization of Novel Transcripts in Pseudorabies Virus. Viruses. 2015;7:2727–44.
182. Metzker ML. Sequencing technologies — the next generation. Nat Rev Genet. 2009;11(1):31–46.
183. Mbala-kingebeni P, Mukadi D, Mukadi P, Kumakamba C. Rapid Confirmation of the Zaire Ebola Virus in the Outbreak of the Equateur Province in the Democratic Republic of Congo : Implications for Public Health Interventions. Clin Infect Dis. 2019;68(2):330–3.
184. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc. 2017;12(6).
185. Sequence G, Isolate S-, Using O, Nanopore O. Genome Sequence of SARS-CoV-2 Isolate Cali-01, from Colombia, Obtained Using Oxford Nanopore MinION Sequencing. Microbiol Resour Announc. 2020;9:e00573-20.
186. Leggett RM, Clark MD. A world of opportunities with nanopore sequencing. Jpurnal Exp Bot. 2017;68(20):5419–29.
187. Wooley JC, Ye Y. Metagenomics : Facts and Artifacts , and Computational Challenges ã. J Comput Sci Technol. 2010;25(1):71–81.
188. Delwart EL. Viral metagenomics. Rev Med Virol. 2007;17:115–31.
189. Kumar A, Murthy S, Kapoor A. Evolution of selective-sequencing approaches for virus discovery and virome analysis. Virus Res. 2017;239:172–9.
190. Li L, Deng X, Mee ET, Collot-Teixeira S, Anderson R, Schepelmann S, et al. Comparing

viral metagenomics methods using a highly multiplexed human viral pathogens reagent. J Virol Methods. 2014 Dec 11;213C:139–46.
191. Calvet G, Aguiar RS, Melo ASO, Sampaio SA, de Filippis I, Fabri A, et al. Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. Lancet Infect Dis. 2016 Jun 1;16(6):653–60.
192. Halary S, Duraisamy R, Fancello L, Monteil-Bouchard S, Jardot P, Biagini P, et al. Novel single-stranded DNA circular viruses in pericardial fluid of patient with recurrent pericarditis. Vol. 22, Emerging Infectious Diseases. Centers for Disease Control and Prevention (CDC); 2016. p. 1839–41.
193. Allander T, Emerson SU, Engle RE, Purcell RH, Bukh J. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. Proc Natl Acad Sci U S A. 2001;98(20):11609–14.
194. Cann AJ, Fandrich SE, Heaphy S. Analysis of the Virus Population Present in Equine Faeces Indicates the Presence of Hundreds of Uncharacterized Virus Genomes w. Virus Genes. 2005;30(2):151–6.
195. Zhang T, Breitbart M, Lee WH, Run J, Wei CL, Wee S, et al. RNA Viral Community in Human Feces : Prevalence of Plant Pathogenic Viruses. Plos Biol. 2006;4(1):e3.
196. Matsui SM, Oshiro LS, Reyes GR. The isolation and characterization of a Norwalk virus-specific cDNA. J Clin Invest. 1991;87(4):1456–61.
197. Matsui SM, Kim JP, Greenberg HB, Young LAVM, Smith LS, Lewis TL, et al. Cloning and Characterization of Human Astrovirus Immunoreactive Epitopes. J Virol. 1993;67(3):1712–5.
198. Jones MS, Kapoor A, Lukashov V V, Hecht F, Delwart E, Jones MS, et al. New DNA Viruses Identified in Patients with Acute Viral Infection Syndrome New DNA Viruses Identified in Patients with Acute Viral Infection Syndrome. J Virol. 2005;79(13):8230–6.
199. Tammi MT, Eriksson M, Allander T, Tammi MT, Eriksson M, Bjerkner A, et al. Cloning of a human parvovirus by molecular screening of respiratory tract samples. PNAS. 2005;102(36):12891–6.
200. Phan TG, Li L, O'Ryan MG, Cortes H, Mamani N, Bonkoungou IJO, et al. A third gyrovirus species in human faeces. J Gen Virol. 2012 Jun;93(Pt 6):1356–61.
201. Xu B, Liu L, Huang X, Ma H, Zhang Y, Du Y, et al. Metagenomic Analysis of Fever, Thrombocytopenia and Leukopenia Syndrome (FTLS) in Henan Province, China: Discovery of a New Bunyavirus. Palacios G, editor. PLoS Pathog. 2011 Nov 17;7(11):e1002369.
202. Breitbart M, Rohwer F. Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. Biotechniques. 2005;39(5):729–36.
203. Tanaka Y, Primi D, Wang RYH, Umemura T, Yeo AET, Mizokami M, et al. Genomic and Molecular Evolutionary Analysis of a Newly Identified Infectious Agent (SEN Virus) and Its Relationship to the TT Virus Family. J Infect Dis. 2001;183:359–67.
204. Flygare S, Simmon K, Miller C, Qiao Y, Kennedy B, Di Sera T, et al. Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling. Genome Biol. 2016;17(1):111.
205. Wood DE, Salzberg SL. Kraken : ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15(R46).
206. Kalantar KL, Carvalho T, Bourcy CFA De, Dimitrov B, Dingle G, Egger R, et al. IDseq — An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. Gigascience. 2020;9:1–14.
207. Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D, Samayoa E, et al. A

cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. Genome Res. 2014;24:1180–92.
208. Nooij S, Schmitz D, Vennema H, Kroneman A. Overview of Virus Metagenomic Classification Methods and Their Biological Applications. Front Microbiol. 2018;9:749.
209. Grumaz S, Grumaz C, Vainshtein Y, Stevens P, Glanz K, Decker SO, et al. Enhanced Performance of Next-Generation Sequencing Diagnostics Compared With Standard of Care Microbiological Diagnostics in Patients Suffering From Septic Shock. Crit Care Med. 2019;47(5):e394–402.
210. Li S, Jiang W, Peng J, Du B, Weng L. Herpes simplex virus associated sepsis in an immunocompetent adult : the value of next-generation sequencing. Chin Med J (Engl). 2020;133(14):1727–8.
211. Rozo M, Id KLS, Philipson C, Fitkariwala A, Nhim D, Som T, et al. An Observational Study of Sepsis in Takeo Province Cambodia : An in-depth examination of pathogens causing severe infections. PLoS Negl Trop Dis. 2020;14(8):e0008381.
212. Long Y, Zhang Y, Gong Y, Sun R, Su L, Lin X. Diagnosis of Sepsis with Cell-free DNA by Next-Generation Sequencing Technology in ICU Patients. Arch Med Res. 2016;47(5):365–71.
213. Saha S, Ramesh A, Kalantar K, Malaker R, Hasanuzzaman M, Khan LM, et al. Unbiased Metagenomic Sequencing for Pediatric Meningitis in Bangladesh Reveals Neuroinvasive Chikungunya Virus Outbreak and Other Unrealized Pathogens. MBio. 2019 Dec 17;10(6):e02877-19.
214. Miller S, Naccache SN, Samayoa E, Messacar K, Arevalo S, Federman S, et al. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. Genome Res. 2019 May 1;29(5).
215. Naccache SN, Peggs KS, Mattes FM, Phadke R, Garson JA, Grant P, et al. Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing. Clin Infect Dis. 2015 Mar 15;60(6):919–23.
216. Zanella M, Lenggenhager L, Schrenzel J, Cordey S, Kaiser L. High-throughput sequencing for the aetiologic identification of viral encephalitis, meningoencephalitis, and meningitis. A narrative review and. Clin Microbiol Infect. 2019;25:422–30.
217. Anthony SJ, Epstein JH, Murray KA, Navarrete-macias I, Zambrana-torrelio CM, Solovyov A, et al. A Strategy To Estimate Unknown Viral Diversity in Mammals. MBio. 2013;4(5):e00598-13.
218. Zhou P, Yang X Lou, Wang XJ, Hu B, Zhang L. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. bioRxiv. 2020;21(1):1–9.
219. Allen T, Murray KA, Zambrana-torrelio C, Morse SS, Rondinini C, Di Marco M, et al. Global hotspots and correlates of emerging zoonotic diseases. Nat Commun. 2017 Dec 24;8(1):1124.
220. Endoh D, Mizutani T, Kirisawa R, Maki Y, Saito H, Kon Y, et al. Species-independent detection of RNA virus by representational difference analysis using non-ribosomal hexanucleotides for reverse transcription. Nucleic Acids Res. 2005;33(6):1–11.
221. Nguyen AT, Tran TT, Hoang VMT, Nghiem NM, Le NNTNNT, Le TTMT Van, et al. Development and evaluation of a non-ribosomal random PCR and next-generation sequencing based assay for detection and sequencing of hand, foot and mouth disease pathogens. Virol J. 2016;13(1):125.
222. Ngoc C Le, Tran T, Thanh T, Tran P, Lan T, Mai TN, et al. Differential prevalence and geographic distribution of hepatitis C virus genotypes in acute and chronic hepatitis C patients in. PLoS One. 2019;14(3):e0212734.

223. Tan L Van, Thai LH, Phu NH, Dang H, Nghia T, Chuong L Van. Viral Aetiology of Central Nervous System Infections in Adults Admitted to a Tertiary Referral Hospital in Southern Vietnam over 12 Years. PLoS Negl Trop Dis. 2014;8(8):e3127.
224. Froussard P. A random-PCR method (rPCR) to construct whole cDNA library from low amounts of RNA. Nucleic Acids Res. 1992;20(11):2900.
225. van Tan L, van Doorn HR, van der Hoek L, Hien VM, Jebbink MF, Ha DQ, et al. Random PCR and ultracentrifugation increases sensitivity and throughput of VIDISCA for screening of pathogens in clinical specimens. J Infect Dev Ctries. 2011;5(2):142–8.
226. Graf EH, Simmon KE, Tardif KD, Hymas W, Flygare S, Eilbeck K, et al. Unbiased Detection of Respiratory Viruses Using RNA-seq-Based Metagenomics: A Systematic Comparison to A Commercial PCR Panel. J Clin Microbiol. 2016;54(January):JCM.03060-15.
227. Lewandowska DW, Zagordi O, Geissberger F-D, Kufner V, Schmutz S, Böni J, et al. Optimization and validation of sample preparation for metagenomic sequencing of viruses in clinical samples. Microbiome. 2017;5(1):94.
228. Goldstein B, Giroir B, Randolph A. International pediatric sepsis consensus conference: Definitions for sepsis and organ dysfunction in pediatrics*. Pediatr Crit Care Med. 2005;6(1):2–8.
229. Aiemjoy K, Altan E, Aragie S, Fry DM, Phan TG, Deng X, et al. Viral species richness and composition in young children with loose or watery stool in Ethiopia. BMC Infect Dis. 2019;19(1):1–10.
230. Hummel KB, Lowe L, Bellini WJ, Rota PA. Development of quantitative gene-specific real-time RT-PCR assays for the detection of measles virus in clinical specimens. J Virol Methods. 2006;132(1–2):166–73.
231. Dung TTN, Phat VV, Nga TVTT, My PVT, Duy PT, Campbell JI, et al. The validation and utility of a quantitative one-step multiplex RT real-time PCR targeting Rotavirus A and Norovirus. J Virol Methods. 2013;187(1):138–43.
232. Thanh T, Anh N, Tham N, Van H, Sabanathan S, Qui P, et al. Validation and utilization of an internally controlled multiplex Real-time RT-PCR assay for simultaneous detection of enteroviruses and enterovirus A71 associated with hand foot and mouth disease. Virol J. 2015;12(1):85.
233. Frentiu FD, Zakir T, Walker T, Popovici J, Pyke AT, van den Hurk A, et al. Limited Dengue Virus Replication in Field-Collected Aedes aegypti Mosquitoes Infected with Wolbachia. PLoS Negl Trop Dis. 2014;8(2):1–10.
234. Hue KDT, Tuan TV, Thi HTN, Bich CTN, Anh HH Le, Wills BA, et al. Validation of an internally controlled one-step real-time multiplex RT-PCR assay for the detection and quantitation of dengue virus RNA in plasma. J Virol Methods. 2011;177(2):168–73.
235. Chook JB, Ong LY, Takebe Y, Chan KG, Choo M, Kamarulzaman A, et al. Molecular detection of HIV-1 subtype B, CRF01-AE, CRF33-01B, and newly emerging recombinant lineages in Malaysia. Am J Trop Med Hyg. 2015;92(3):507–12.
236. Jansen RR, Schinkel J, Koekkoek S, Pajkrt D, Beld M, Jong MD d. de, et al. Development and evaluation of a four-tube real time multiplex PCR assay covering fourteen respiratory viruses, and comparison to its corresponding single target counterparts. J Clin Virol. 2011;51(3):179–85.
237. Harvala H, Robertson I, McWilliam Leitch EC, Benschop K, Wolthers KC, Templeton K, et al. Epidemiology and clinical associations of human parechovirus respiratory infections. J Clin Microbiol. 2008;46(10):3446–53.
238. Drexler JF, De Souza Luna LK, Stöcker A, Silva Almeida P, Medrado Ribeiro TC, Petersen N, et al. Circulation of 3 lineages of a novel saffold cardiovirus in humans.

Emerg Infect Dis. 2008;14(9):1398–405.
239. Haramoto E, Kitajima M, Otagiri M. Development of a reverse transcription-quantitative PCR assay for detection of salivirus/klassevirus. Appl Environ Microbiol. 2013;79(11):3529–32.
240. Kuypers J, Campbell AP, Guthrie KA, Wright NL, Englund JA, Corey L, et al. WU and KI polyomaviruses in respiratory samples from allogeneic hematopoietic cell transplant recipients. Emerg Infect Dis. 2012;18(10):1580–8.
241. Tanaka N, Kimura H, Hoshino Y, Kato K, Yoshikawa T, Asano Y, et al. Monitoring four herpesviruses in unrelated cord blood transplantation. Bone Marrow Transplant. 2000;26(11):1193–7.
242. Boom R, Sol CJA, Schuurman T, Breda A Van, Weel JFL, Beld M, et al. Human Cytomegalovirus DNA in Plasma and Serum Specimens of Renal Transplant Recipients Is Highly Fragmented. J Clin Microbiol. 2002;40(11):4105–13.
243. Phan TG, Mori D, Deng X, Rajidrajith S, Ranawaka U, Fan Ng TF, et al. Small viral genomes in unexplained cases of human encephalitis, diarrhea, and in untreated sewage. Virology. 2015;482:98–104.
244. Tan L Van, Doorn HR Van, Trung D, Hong T, Phuong T, Vries M De. Identification of a New Cyclovirus in Cerebrospinal Fluid of Patients with Acute Central Nervus System Infections. MBio. 2013;4(3):1–10.
245. Frankel M, Forberg K, Coller KE, Berg MG, Hackett J, Cloherty G, et al. Development of a high-throughput multiplexed real time RT-PCR assay for detection of human pegivirus 1 and 2. J Virol Methods. 2017;241:34–40.
246. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74.
247. Obonyo NG, Schlapbach LJ, Fraser JF. Sepsis: Changing Definitions, Unchanging Treatment. Front Pediatr. 2019;6(January):1–5.
248. Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, et al. The Perils of Pathogen Discovery: Origin of a Novel Parvovirus-Like Hybrid Genome Traced to Nucleic Acid Extraction Spin Columns. J Virol. 2013;87(22):11966–77.
249. Brito F, Cordey S, Delwart E, Deng X, Tirefort D, Lemoine-Chaduc C, et al. Metagenomics analysis of the virome of 300 concentrates from a Swiss platelet bank. Vox Sang. 2018;113(6):601–4.
250. Nguyen TA, Le NTN, Hoang MT Van, Nguyen TTH, Tran TT, Vu TTH, et al. Emerging Coxsackievirus A6 causing Hand, Foot and Mouth Disease, Vietnam. Emerg Infect Dis. 2018;24(4):17–9.
251. Nguyen TTCB, Ngo NQM, Phan TQ, Tran THC, Ho DTN, Lien AH Do, et al. Enterovirus serotypes in patients with central nervous system and respiratory infections in Viet Nam 1997 – 2010. Virol J. 2018;15(69):1–8.
252. Honkanen H, Oikarinen S, Pakkanen O, Ruokoranta T, Pulkki MM, Laitinen OH, et al. Human enterovirus 71 strains in the background population and in hospital patients in Finland. J Clin Virol. 2013;56(4):348–53.
253. Kong KL, Lau JSY, Goh SM, Wilson HL, Catton M, Korman TM. Myocarditis caused by human parechovirus in adult. Emerg Infect Dis. 2017;23(9):1571–3.
254. Ooi MH, Wong SC, Lewthwaite P, Cardosa MJ, Solomon T. Clinical features, diagnosis, and management of enterovirus 71. Lancet Neurol. 2010;9(11):1097–105.
255. Nhan LNT, Hong NTT, Nhu LNT, Nguyet LA, Ny NTH, Thanh TT. Severe enterovirus A71 associated hand , foot and mouth disease , Vietnam , 2018 : preliminary report of an impending outbreak. Eurosurveillance. 2018;23(46):1–5.

256. Tan SZK, Tan MZY, Prabakaran M. Saffold virus, an emerging human cardiovirus. Rev Med Virol. 2017;27(1):1–11.
257. Nielsen ACY, Böttiger B, Banner J, Hoffmann T, Nielsen LP. Serious invasive Saffold virus infections in children, 2009. Emerg Infect Dis. 2012;18(1):7–12.
258. Reuter G, Pankovics P, Boros Á. Saliviruses—the first knowledge about a newly discovered human picornavirus. Rev Med Virol. 2017;27(1):1–10.
259. Nielsen TS, Nielsen AY, Banner J, Hansen J, Baandrup U, Nielsen LP. Saffold virus infection associated with human myocarditis. J Clin Virol. 2016;74:78–81.
260. Himeda T, Hosomi T, Asif N, Shimizu H, Okuwa T, Muraki Y, et al. The preparation of an infectious full-length cDNA clone of Saffold virus. Virol J. 2011;8(1):110.
261. Hertzler S, Liang Z, Treso B, Lipton HL. Adaptation of Saffold Virus 2 for High-Titer Growth in Mammalian Cells. J Virol. 2011;85(14):7411–8.
262. Tan SZK, Chua KB, Xu Y, Prabakaran M. The pathogenesis of saffold virus in AG129 mice and the effects of its truncated L protein in the central nervous system. Viruses. 2016;8(24).
263. Yang J, Yang F, Ren L, Xiong Z, Wu Z, Dong J, et al. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. J Clin Microbiol. 2011 Oct 1;49(10):3463–9.
264. Wilson MR, Fedewa G, Stenglein MD, Olejnik J, Rennick LJ, Nambulli S, et al. Multiplexed Metagenomic Deep Sequencing To Analyze the Composition of High-Priority Pathogen Reagents. mSystems. 2016;1(4):e00058-16.
265. Mitra A, Skrzypczak M, Ginalski K, Rowicka M. Strategies for achieving high sequencing accuracy for low diversity samples and avoiding sample bleeding using Illumina platform. PLoS One. 2015;10(4):1–21.
266. Berg MG, Lee D, Coller K, Frankel M, Aronsohn A, Cheng K, et al. Discovery of a Novel Human Pegivirus in Blood Associated with Hepatitis C Virus Co-Infection. PLoS Pathog. 2015 Dec;11(12):e1005325.
267. Kapoor A, Kumar A, Simmonds P, Bhuva N, Chauhan LS, Lee B, et al. Virome analysis of transfusion recipients reveals a novel human virus that shares genomic features with hepaciviruses and pegiviruses. MBio. 2015;6(5):1–12.
268. Wang H, Wan Z, Xu R, Guan Y, Zhu N, Li J, et al. A Novel Human Pegivirus, HPgV-2 (HHpgV-1), Is Tightly Associated With Hepatitis C Virus (HCV) Infection and HCV/Human Immunodeficiency Virus Type 1 Coinfection. Clin Infect Dis. 2018;66(1):29–35.
269. Kandathil AJ, Breitwieser FP, Sachithanandham J, Robinson M, Mehta SH, Timp W, et al. Presence of human hepegivirus-1 in a cohort of people who inject drugs. Ann Intern Med. 2017 Jul 4;167(1):1–7.
270. Wang H, Wan Z, Sun Q, Zhu N, Li T, Ren X, et al. Second Human Pegivirus in Hepatitis C virus-infected and Hepatitis C virus/HIV-1 co-infected persons who injected drugs, China. Emerg Infect Dis. 2018;24(5):0–3.
271. Fattovich G, Boscaro S, Noventa F, Pornaro E, Stenico D, Alberti a, et al. Influence of hepatitis delta virus infection on progression to cirrhosis in chronic hepatitis type B. J Infect Dis. 1987;155(5):931–5.
272. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28(10):2731–9.
273. Bijvand Y, Aghasadeghi MR, Sakhaee F, Pakzad P, Vaziri F, Saraji AA, et al. First detection of human hepegivirus-1 (HHpgV-1) in Iranian patients with hemophilia. Sci Rep. 2018;8(1):8–13.

274. Rodgers MA, Holzmayer V, Vallari A, Olivo A, Forberg K, Fuhrman J, et al. Hepatitis C virus surveillance and identification of human pegivirus 2 in a large Cameroonian cohort. J Viral Hepat. 2019;26(1):30–7.
275. Barclay ST, Cooke GS, Holtham E, Gauthier A, Schwarzbard J, Atanasov P, et al. A new paradigm evaluating cost per cure of HCV infection in the UK. Hepatol Med Policy. 2016 Apr;1(1):2.
276. Ireland G, Delpech V, Kirwan P, Croxford S, Lattimore S, Sabin C, et al. Prevalence of diagnosed HIV infection among persons with hepatitis C virus infection: England, 2008-2014. HIV Med. 2018;1–8.
277. Rabaa MA, Tue NT, Phuc TM, Carrique-Mas J, Saylors K, Cotten M, et al. The Vietnam Initiative on Zoonotic Infections (VIZIONS): A Strategic Approach to Studying Emerging Zoonotic Infectious Diseases. Ecohealth. 2015;12(4):726–35.
278. Anh NT, Hong NTT, Nhu LNT, Thanh TT, Lau C-Y, Limmathurotsakul D, et al. Viruses in Vietnamese Patients Presenting with Community-Acquired Sepsis of Unknown Cause. J Clin Microbiol. 2019 Sep 1;57(9):e00386-19.
279. Hošnjak L, Kocjan BJ, Kušar B, Seme K, Poljak M. Rapid detection and typing of Molluscum contagiosum virus by FRET-based real-time PCR. J Virol Methods. 2013;187(2):431–4.
280. Kroneman A, Vennema H, Deforche K, Avoort H, Peñaranda S, Oberste MS, et al. An automated genotyping tool for enteroviruses and noroviruses. J Clin Virol. 2011 Jun;51(2):121–5.
281. Sun Y, Miao Z, Yan J, Gong L, Chen Y, Chen Y, et al. Sero-molecular epidemiology of enterovirus-associated encephalitis in Zhejiang Province , China , from 2014 to 2017. Int J Infect Dis. 2019;79:58–64.
282. Richter J, Tryfonos C, Christodoulou C. Molecular epidemiology of enteroviruses in Cyprus 2008-2017. PLoS One. 2019;1–18.
283. Lema C, Torres C, Van der Sanden S, Cisterna D, Freire MC, Gómez RM. Global phylodynamics of Echovirus 30 revealed differential behavior among viral lineages. Virology. 2019;531(February):79–92.
284. Schibler M, Brito F, Zanella MC, Zdobnov EM, Laubscher F, L'Huillier AG, et al. Viral Sequences Detection by High-Throughput Sequencing in Cerebrospinal Fluid of Individuals with and without Central Nervous System Disease. Genes (Basel). 2019;10(8):1–12.
285. Zayyad Z, Spudich S. Neuropathogenesis of HIV: From Initial Neuroinvasion to HIV-Associated Neurocognitive Disorder (HAND). Curr HIV/AIDS Rep. 2015;12(1):16–24.
286. Zhou C, Zhang S, Gong Q, Hao A. A novel gemycircularvirus in an unexplained case of child encephalitis. Virol J. 2015;12(1):197.
287. Wilson MR, Sample HA, Zorn KC, Arevalo S, Yu G, Neuhaus J, et al. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. N Engl J Med. 2019;380(24):2327–40.
288. Sinha R, Stanley G, Gulati GS, Ezran C, Travaglini KJ, Wei E, et al. Index switching causes "spreading-of-signal" among multiplexed samples in Illumina HiSeq 4000 DNA sequencing. bioRxiv. 2017 Apr 9;125724.
289. Nguyen TTHTA, Nguyen TTHTA, Nguyen Thi Hoang M, Ho DTN, Le NNT, Tran tan T, et al. Performance of metagenomic next-generation sequencing for the diagnosis of viral meningoencephalitis in a resource limited setting. Open Forum Infect Dis. 2020;ofaa046.
290. Asplund M, Kjartansdóttir KR, Mollerup S, Vinner L, Fridholm H, Herrera JAR, et al. Contaminating viral sequences in high-throughput sequencing viromics: a linkage study

of 700 sequencing libraries. Clin Microbiol Infect. 2019;25(10):1277–85.
291. Holmes EC. Reagent contamination in viromics : all that glitters is not gold. Clin Microbiol Infect. 2019;25(10):1167–8.
292. Nguyen Thi Hoang M, Nguyen Hoan P, Le Van T, McBride A, Ho Dang Trung N, Tran Tan T, et al. First reported cases of anti-NMDA receptor encephalitis in Vietnamese adolescents and adults. J Neurol Sci. 2017;373:250–3.
293. Edridge AWD, Deijs M, Van Zeggeren IE, Kinsella CM, Jebbink MF, Bakker M, et al. Viral metagenomics on cerebrospinal fluid. Genes (Basel). 2019;10(5).
294. Uchida K, Shinohara M, Shimada SI, Segawa Y, Doi R, Gotoh A, et al. Rapid and sensitive detection of mumps virus RNA directly from clinical samples by real-time PCR. J Med Virol. 2005;75(3):470–4.
295. Heemskerk AD, Donovan J, Thu DDA, Marais S, Chaidir L, Dung VTM, et al. Improving the microbiological diagnosis of tuberculous meningitis: A prospective, international, multicentre comparison of conventional and modified Ziehl–Neelsen stain, GeneXpert, and culture of cerebrospinal fluid. J Infect. 2018;77(6):509–15.
296. Zhi N, Hu G, Wan Z, Zheng X, Liu X, Wong S, et al. Erratum: Hybrid DNA virus in Chinese patients with seronegative hepatitis discovered by deep sequencing (Proceedings of the National Academy of Sciences of the United States of America (2013) 110 (10264-10269)). Proc Natl Acad Sci U S A. 2014;111(11):4344–5.
297. Chiu CY, Miller SA. Clinical metagenomics. Nat Rev Genet. 2019 Jun 27;20(6):341–55.
298. Sekeyová Z, Danchenko M, Filipčík P, Fournier PE. Rickettsial infections of the central nervous system. PLoS Negl Trop Dis. 2019;13(8):1–18.
299. Huang GKL, Tio SY, Caly L, Nicholson S, Thevarajan I, Papadakis G, et al. Prolonged Detection of Japanese Encephalitis Virus in Urine and Whole Blood in a Returned Short-term Traveler. Open Forum Infect Dis. 2017;4(4):4–6.
300. Mai NTH, Phu NH, Nhu LNT, Hong NTT, Hanh NHH, Nguyet LA, et al. Central Nervous System Infection Diagnosis by Next-Generation Sequencing: A Glimpse Into the Future? Open forum Infect Dis. 2017;4(2):ofx046.
301. Lim ES, Zhou Y, Zhao G, Bauer IK, Droit L, Ndao IM, et al. Early life dynamics of the human gut virome and bacterial microbiome in infants. Nat Med. 2015;21(10):1228–34.
302. Wylie KM, Mihindukulasuriya KA, Sodergren E, Weinstock GM, Storch GA. Sequence analysis of the human virome in febrile and afebrile children. PLoS One. 2012;7(6):e27735.
303. Thorburn F, Bennett S, Modha S, Murdoch D, Gunson R, Murcia PR. The use of next generation sequencing in the diagnosis and typing of respiratory infections. J Clin Virol. 2015;69:96–100.
304. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265–9.
305. McCabe L, White IR, Chau NVV, Barnes E, Pett SL, Cooke GS, et al. The design and statistical aspects of VIETNARMS: A strategic post-licensing trial of multiple oral direct-acting antiviral hepatitis C treatment strategies in Vietnam. Trials. 2020;21(1):1–12.
306. Thaipadunpanit J, Chierakul W, Wuthiekanun V, Limmathurotsakul D, Amornchai P, Boonslip S, et al. Diagnostic accuracy of real-time PCR assays targeting 16S rRNA and lipl32 genes for human leptospirosis in Thailand: A case-control study. PLoS One. 2011;6(1):1–6.
307. Laras K, Cao bao van, Bounlu K, Nguyen TKT, Olson JG, Thongchanh S, et al. The importance of tourism in South-East Asia. Am J Trop Med Hyg. 2002;67(3):278–86.
308. Blacksell SD, Jenjaroen K, Phetsouvanh R, Wuthiekanun V, Day NPJ, Newton PN, et al. Accuracy of AccessBio Immunoglobulin M and Total Antibody Rapid

Immunochromatographic Assays for the Diagnosis of Acute Scrub Typhus Infection. Clin Vaccine Immunol. 2010 Feb 1;17(2):263–6.
309. Jiang J, Chan T-C, Temenak JJ, Dasch GA, Ching W-M, Richards AL. Development of a quantitative real-time polymerase chain reaction assay specific for Orientia tsutsugamushi. Am J Trop Med Hyg. 2004 Apr;70(4):351–6.
310. Henry KM, Jiang J, Rozmajzl PJ, Azad AF, Macaluso KR, Richards AL. Development of quantitative real-time PCR assays to detect Rickettsia typhi and Rickettsia felis, the causative agents of murine typhus and flea-borne spotted fever. Mol Cell Probes. 2007 Feb;21(1):17–23.
311. Cherkaoui A, Emonet S, Ceroni D, Candolfi B, Hibbs J, Francois P, et al. Development and validation of a modified broad-range 16S rDNA PCR for diagnostic purposes in clinical microbiology. J Microbiol Methods. 2009 Nov;79(2):227–31.
312. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol. 1991 Jan;173(2):697–703.
313. Klempa B, Fichet-Calvet E, Lecompte E, Auste B, Aniskin V, Meisel H, et al. Hantavirus in African wood mouse, Guinea. Emerg Infect Dis. 2006 May;12(5):838–40.
314. Cardosa MJ, Wang SM, Sum MSH, Tio PH. Antibodies against prM protein distinguish between previous infection with dengue and Japanese encephalitis viruses. BMC Microbiol. 2002 May 5;2:9.
315. van Doornum GJJ, Guldemeester J, Osterhaus ADME, Niesters HGM. Diagnosing herpesvirus infections by real-time amplification and rapid culture. J Clin Microbiol. 2003 Feb;41(2):576–80.
316. de Jong MD, Weel JF, Schuurman T, Wertheim-van Dillen PM, Boom R. Quantitation of varicella-zoster virus DNA in whole blood, plasma, and serum by PCR and electrochemiluminescence. J Clin Microbiol. 2000 Jul;38(7):2568–73.
317. Beld M, Minnaar R, Weel J, Sol C, Damen M, van der Avoort H, et al. Highly sensitive assay for detection of enterovirus in clinical specimens by reverse transcription-PCR with an armored RNA internal control. J Clin Microbiol. 2004 Jul 1;42(7):3059–64.
318. Corless CE, Guiver M, Borrow R, Edwards-Jones V, Fox AJ, Kaczmarski EB. Simultaneous detection of Neisseria meningitidis, Haemophilus influenzae, and Streptococcus pneumoniae in suspected cases of meningitis and septicemia using real-time PCR. J Clin Microbiol. 2001 Apr 1;39(4):1553–8.
319. Nga TVT, Nghia HDT, Tu LTP, Diep TS, Mai NTH, Chau TTH, et al. Real-time PCR for detection of Streptococcus suis serotype 2 in cerebrospinal fluid of human patients with meningitis. Diagn Microbiol Infect Dis. 2011 Aug;70(4):461–7.
320. Heddema ER, Beld MGHM, de Wever B, Langerak AAJ, Pannekoek Y, Duim B. Development of an internally controlled real-time PCR assay for detection of Chlamydophila psittaci in the LightCycler 2.0 system. Clin Microbiol Infect. 2006 Jun;12(6):571–5.
321. Pitcher D, Chalker VJ, Sheppard C, George RC, Harrison TG. Real-time detection of Mycoplasma pneumoniae in respiratory samples with an internal processing control. J Med Microbiol. 2006 Feb 1;55(Pt 2):149–55.
322. Reischl U, Lehn N, Sanden GN, Loeffelholz MJ. Real-time PCR assay targeting IS481 of Bordetella pertussis and molecular basis for detecting Bordetella holmesii. J Clin Microbiol. 2001 May 1;39(5):1963–6.
323. Logan C, O'Leary JJ, O'Sullivan N. Real-time reverse transcription-PCR for detection of rotavirus and adenovirus as causative agents of acute viral gastroenteritis in children. J Clin Microbiol. 2006 Sep;44(9):3189–95.
324. Logan C, O'Leary JJ, O'Sullivan N. Real-time reverse transcription PCR detection of

norovirus, sapovirus and astrovirus as causative agents of acute viral gastroenteritis. J Virol Methods. 2007;146(1–2):36–44.
325. Ursu K, Harrach B, Matiz K, Benko M. DNA sequencing and analysis of the right-hand part of the genome of the unique bovine adenovirus type 10. J Gen Virol. 2004;85(3):593–601.
326. Woo PCY, Lau SKP, Lam CSF, Lai KKY, Huang Y, Lee P, et al. Comparative Analysis of Complete Genome Sequences of Three Avian Coronaviruses Reveals a Novel Group 3c Coronavirus. J Virol. 2009;83(2):908–17.
327. Li L, Kapoor A, Slikas B, Bamidele OS, Wang C, Shaukat S, et al. Multiple Diverse Circoviruses Infect Farm Animals and Are Commonly Found in Human and Chimpanzee Feces. J Virol. 2010;84(4):1674–82.
328. Bailey L, Newman JF., Porterfield JS. The Multiplication of Nodamura Virus in Insect and Mammalian Cell Cultures. J Gen Virol. 1966;26:15–20.
329. Hansen TA, Mollerup S, Nguyen NP, White NE, Coghlan M, Alquezar-Planas DE, et al. High diversity of picornaviruses in rats from different continents revealed by deep sequencing. Emerg Microbes Infect. 2016;5(8):e90.
330. Woo PCY, Lau SKP, Teng JLL, Tsang AKL, Joseph M, Wong EYM, et al. Metagenomic analysis of viromes of dromedary camel fecal samples reveals large number and high diversity of circoviruses and picobirnaviruses. Virology. 2014;471–473:117–25.
331. Woo PCY, Lau SKP, Bai R, Teng JLL, Lee P, Martelli P, et al. Complete Genome Sequence of a Novel Picobirnavirus, Otarine Picobirnavirus, Discovered in California Sea Lions. J Virol. 2012;86(11):6377–8.
332. Wu Z, Yang L, Ren X, He G, Zhang J, Yang J, et al. Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. ISME J. 2016;10(3):609–20.
333. Xiao P, Li C, Zhang Y, Han J, Guo X, Xie L, et al. Metagenomic Sequencing From Mosquitoes in China Reveals a Variety of Insect and Human Viruses. Front Cell Infect Microbiol. 2018;8(October):1–11.
334. Phan TG, Mendoza J del V, Sadeghi M, Altan E, Deng X, Delwart E, et al. Sera of Peruvians with fever of unknown origins include viral nucleic acids from non-vertebrate hosts. Virus Genes. 2018 Feb;54(1):33–40.
335. Ngoi CN, Siqueira J, Li L, Deng X, Mugo P, Graham SM, et al. The plasma virome of febrile adult kenyans shows frequent parvovirus B19 infections and a novel arbovirus (Kadipiro virus). J Gen Virol. 2016;97(12):3359–67.
336. Aguiar ERGR, Olmo RP, Paro S, Ferreira FV, De Faria IJDS, Todjro YMH, et al. Sequence-independent characterization of viruses based on the pattern of viral small RNAs produced by the host. Nucleic Acids Res. 2015;43(13):6191–206.
337. Belaganahalli MN, Maan S, Maan NS, Nomikou K, Pritchard I, Lunt R, et al. Full genome sequencing and genetic characterization of eubenangee viruses identify pata virus as a distinct species within the genus orbivirus. PLoS One. 2012;7(3).
338. Horta AB, Ardisson-Araujo DMP, da Silva LA, de Melo FL, da Silva Morgado F, Franco Lemos MV, et al. Genomic analysis of a cypovirus isolated from the eucalyptus brown looper, Thyrinteina arnobia (Stoll, 1782) (Lepidoptera: Geometridae). Virus Res. 2018;253(2018):62–7.
339. Medeiros DB de A, Diniz Júnior JAP, Cardoso JF, Silva SP, da Silva DEA, de Oliveira LF, et al. Nearly complete genome sequence of curionopolis virus, a culicoides-related rhabdovirus isolated in the brazilian Amazon region. Genome Announc. 2014;2(6):e01158-14.
340. da Costa AC, Moron AF, Forney LJ, Linhares IM, Sabino E, Costa SF, et al. Identification

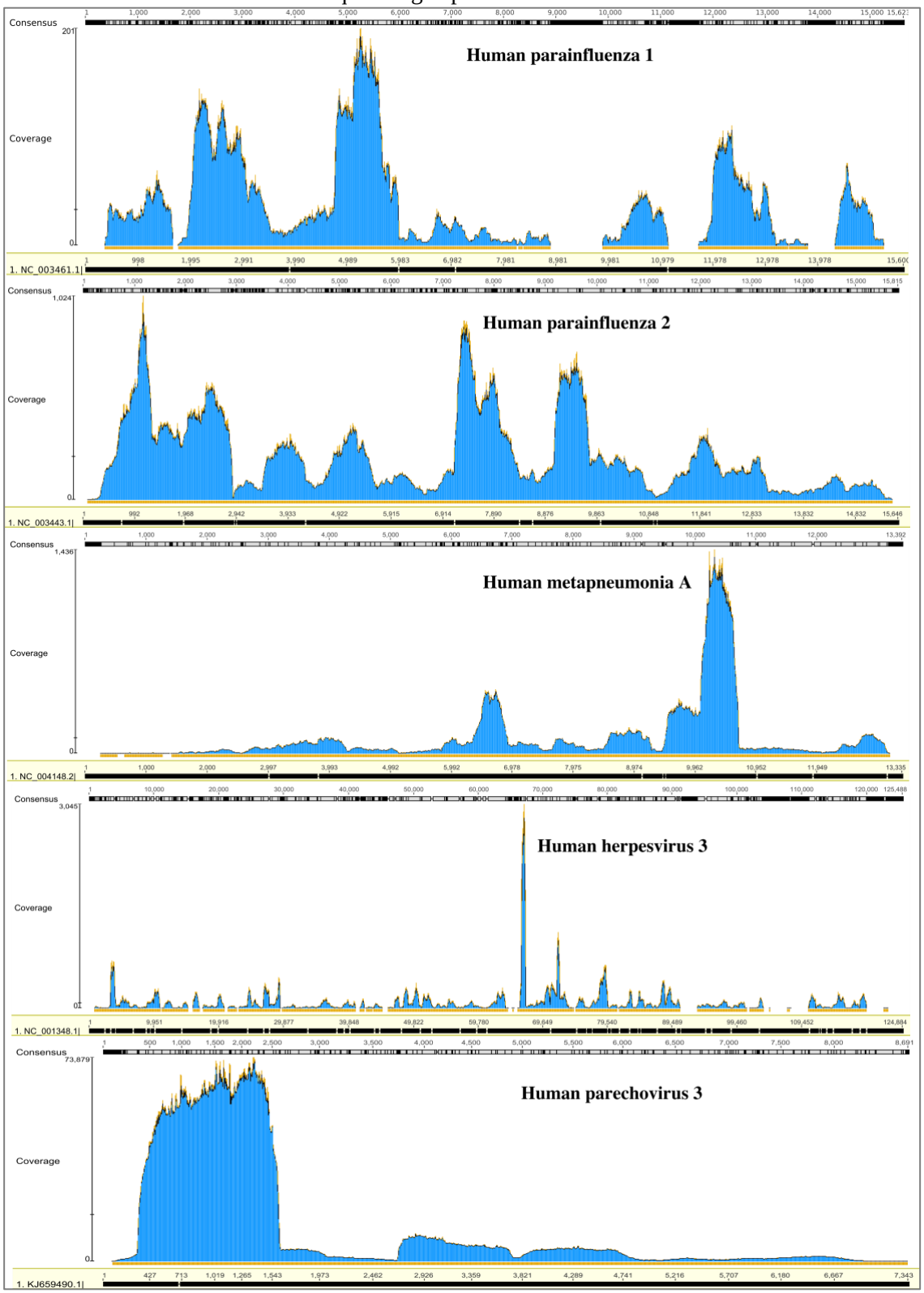
of bacteriophages in the vagina of pregnant women: a descriptive study. BJOG An Int J Obstet Gynaecol. 2020;0–2.
341. Bruenn JA. A closely related group of RNA-dependent RNA polymerases from double-stranded RNA viruses. Nucleic Acids Res. 1993;21(24):5667–9.
342. Taylor DJ, Ballinger MJ, Bowman SM, Bruenn JA. Virus-host co-evolution under a modified nuclear genetic code. PeerJ. 2013;1:e50.
343. Ai YP, Zhong J, Chen CY, Zhu HJ, Gao B Da. A novel single-stranded RNA virus isolated from the rice-pathogenic fungus Magnaporthe oryzae with similarity to members of the family Tombusviridae. Arch Virol. 2016;161(3):725–9.
344. Ng TFF, Marine R, Wang C, Simmonds P, Kapusinszky B, Bodhidatta L, et al. High Variety of Known and New RNA and DNA Viruses of Diverse Origins in Untreated Sewage. J Virol. 2012;86(22):12161–75.

Appendices

Appendix 1: List of 96 non-ribosomal random primers and FR20RV primer sequences

Primer	Sequence of primer (5'-3')	Primer	Sequence of primer (5'-3')	Primer	Sequence of primer (5'-3')
FR20RV	GCCGGAGCTCTGCAGATATC	33	GCCGGAGCTCTGCAGATATCTAGTCG	66	GCCGGAGCTCTGCAGATATCTAACGC
1	GCCGGAGCTCTGCAGATATCGATATC	34	GCCGGAGCTCTGCAGATATCGTAGAC	67	GCCGGAGCTCTGCAGATATCGGTCAT
2	GCCGGAGCTCTGCAGATATCTAGTAT	35	GCCGGAGCTCTGCAGATATCCTATAG	68	GCCGGAGCTCTGCAGATATCCTCATA
3	GCCGGAGCTCTGCAGATATCTATAGT	36	GCCGGAGCTCTGCAGATATCTAGCTA	69	GCCGGAGCTCTGCAGATATCAATTTG
4	GCCGGAGCTCTGCAGATATCTATATA	37	GCCGGAGCTCTGCAGATATCACTACT	70	GCCGGAGCTCTGCAGATATCCTGGTA
5	GCCGGAGCTCTGCAGATATCATACTA	38	GCCGGAGCTCTGCAGATATCTAACGA	71	GCCGGAGCTCTGCAGATATCTTCATG
6	GCCGGAGCTCTGCAGATATCATATAT	39	GCCGGAGCTCTGCAGATATCCGACTA	72	GCCGGAGCTCTGCAGATATCGCGATA
7	GCCGGAGCTCTGCAGATATCGTGCAC	40	GCCGGAGCTCTGCAGATATCTACTAG	73	GCCGGAGCTCTGCAGATATCACTAAG
8	GCCGGAGCTCTGCAGATATCACTATA	41	GCCGGAGCTCTGCAGATATCAGTAGT	74	GCCGGAGCTCTGCAGATATCGCATAC
9	GCCGGAGCTCTGCAGATATCCGTAAT	42	GCCGGAGCTCTGCAGATATCGTTAAC	75	GCCGGAGCTCTGCAGATATCCAATAT
10	GCCGGAGCTCTGCAGATATCCTATAC	43	GCCGGAGCTCTGCAGATATCGTCTAC	76	GCCGGAGCTCTGCAGATATCACCGTA
11	GCCGGAGCTCTGCAGATATCTATACG	44	GCCGGAGCTCTGCAGATATCTACAAG	77	GCCGGAGCTCTGCAGATATCGTGCTA
12	GCCGGAGCTCTGCAGATATCTATGCG	45	GCCGGAGCTCTGCAGATATCTACCAG	78	GCCGGAGCTCTGCAGATATCACGCTA
13	GCCGGAGCTCTGCAGATATCGATACT	46	GCCGGAGCTCTGCAGATATCTGGATT	79	GCCGGAGCTCTGCAGATATCATGTCG
14	GCCGGAGCTCTGCAGATATCCGTATA	47	GCCGGAGCTCTGCAGATATCTCGTTA	80	GCCGGAGCTCTGCAGATATCAGCTTA
15	GCCGGAGCTCTGCAGATATCGTATAG	48	GCCGGAGCTCTGCAGATATCATAGTA	81	GCCGGAGCTCTGCAGATATCCGACAT
16	GCCGGAGCTCTGCAGATATCCGGTTA	49	GCCGGAGCTCTGCAGATATCATAGTC	82	GCCGGAGCTCTGCAGATATCGCTATA
17	GCCGGAGCTCTGCAGATATCAATAGT	50	GCCGGAGCTCTGCAGATATCCTAGTA	83	GCCGGAGCTCTGCAGATATCGCTATG
18	GCCGGAGCTCTGCAGATATCCGCATA	51	GCCGGAGCTCTGCAGATATCGTACTA	84	GCCGGAGCTCTGCAGATATCTGTAAG
19	GCCGGAGCTCTGCAGATATCATTACG	52	GCCGGAGCTCTGCAGATATCTAAGTT	85	GCCGGAGCTCTGCAGATATCAACTTA
20	GCCGGAGCTCTGCAGATATCTTAACA	53	GCCGGAGCTCTGCAGATATCATATCC	86	GCCGGAGCTCTGCAGATATCATAACG
21	GCCGGAGCTCTGCAGATATCAGTATC	54	GCCGGAGCTCTGCAGATATCTCGATA	87	GCCGGAGCTCTGCAGATATCATGTTA
22	GCCGGAGCTCTGCAGATATCTGTTAA	55	GCCGGAGCTCTGCAGATATCGTACCA	88	GCCGGAGCTCTGCAGATATCTGGTAT
23	GCCGGAGCTCTGCAGATATCACTATT	56	GCCGGAGCTCTGCAGATATCGTATCA	89	GCCGGAGCTCTGCAGATATCTGCGTA
24	GCCGGAGCTCTGCAGATATCTAACCG	57	GCCGGAGCTCTGCAGATATCATACTC	90	GCCGGAGCTCTGCAGATATCGGATAT
25	GCCGGAGCTCTGCAGATATCCGATAT	58	GCCGGAGCTCTGCAGATATCACATTA	91	GCCGGAGCTCTGCAGATATCCATAGC
26	GCCGGAGCTCTGCAGATATCGTATAC	59	GCCGGAGCTCTGCAGATATCATATTG	92	GCCGGAGCTCTGCAGATATCCATACT
27	GCCGGAGCTCTGCAGATATCAATCCA	60	GCCGGAGCTCTGCAGATATCCGTCTA	93	GCCGGAGCTCTGCAGATATCCGGATA
28	GCCGGAGCTCTGCAGATATCTAGCAC	61	GCCGGAGCTCTGCAGATATCCTTAGT	94	GCCGGAGCTCTGCAGATATCTTACTA
29	GCCGGAGCTCTGCAGATATCATATCG	62	GCCGGAGCTCTGCAGATATCCTTACA	95	GCCGGAGCTCTGCAGATATCACTCGT
30	GCCGGAGCTCTGCAGATATCAATATT	63	GCCGGAGCTCTGCAGATATCTTATGC	96	GCCGGAGCTCTGCAGATATCTAAGGT
31	GCCGGAGCTCTGCAGATATCTATAGC	64	GCCGGAGCTCTGCAGATATCATACGC		
32	GCCGGAGCTCTGCAGATATCCTTGTA	65	GCCGGAGCTCTGCAGATATCCGCTTA		

Appendix 2: Screen snapshots showing coverage of mapping viral reads recovered by mNGS with pretreatment approach #1. Only viruses with nearly/complete genome sequences were showed. The genome coverage/sequencing depth is indicated by the Y axis and orange lines highlight the sequencing depth of 2 or more.



Appendix 3: Diagnostic work-up carried out as per the study protocol of the original reports of CA sepsis and CNS infection studies.

Pathogens[#, **]	**Assay #1 and samples**	**Assay #2 and samples**	**References**
Leptospirosis	Whole-blood PCR	Microagglutination tests of paired sera	(306,307)
Scrub typhus	Whole-blood PCR	IFA of paired sera	(308,309)
Rickettsiosis and murine typhus	Whole-blood PCR		(309,310)
Murine typhus	IFA of paired sera		(308)
Bacteraemia	Whole-blood PCR		(311,312)
Hantavirus	Serum PCR		(313)
Japanese encephalitis virus	CSF ELISA		(314)
Dengue virus	CSF ELISA		(314)
Mumps	CSF ELISA		Mumps virus (Parotitis) IgM ELISA Kit (IBL International, Germany)
Measles	CSF ELISA		Measles virus IgM micro-capture ELISA (IBL International, Hamburg, Germany)
Rubella	CSF ELISA		Rubella virus IgM micro-capture ELISA (Novatec Immundiagnostica Technologie & Waldpark, Germany)
Dengue	CSF PCR		(234)
Herpes simplex virus 1 and 2	CSF PCR		(315)
Varicella-zoster virus	CSF PCR		(316)
Enterovirus	CSF PCR		(317)
Parechovirus	CSF PCR		(236)
N. meningitidis	CSF PCR		(318)
S. pneumoniae	CSF PCR		(318)
H. influenza type b	CSF PCR		(318)
S. suis	CSF PCR		(319)
Respiratory viruses	Pooled nasal-throat swab PCR		(236)
Respiratory bacteria	Pooled nasal-throat swab PCR		(320–322)
Adenovirus	Stool PCR		(323)
Astrovirus	Stool PCR		(324)
Norovirus	Stool PCR		(231)
Rotavirus	Stool PCR		(231)

Notes:

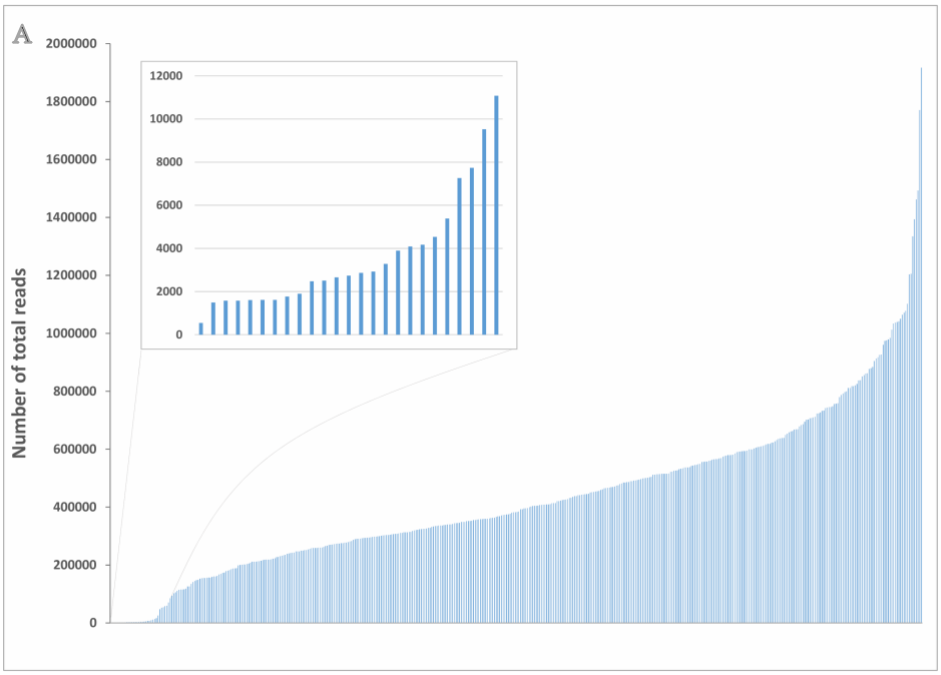
[#]Diagnostic tests performed in every case as part of standard of care at participating hospitals including complete blood count, blood culture, urine culture, gram/ZN smears, and sputum culture if patients have respiratory symptoms, stool examination and stool culture if patients have diarrheal symptoms and CSF examination and CSF culture if patients have neurological symptoms and CNS infection is suspected.

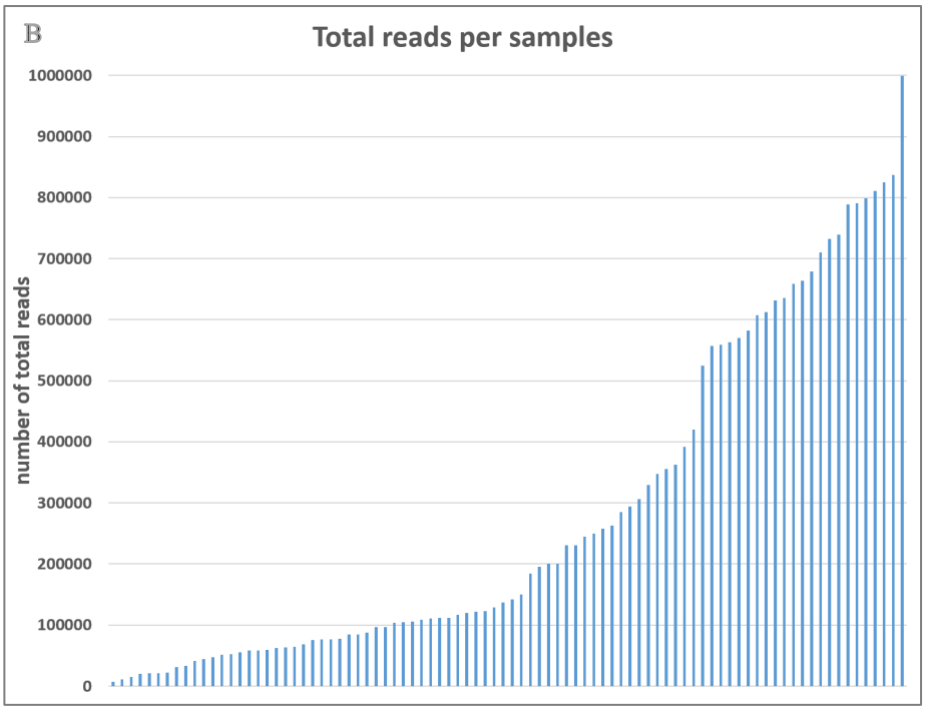
**Diagnostic tests performed in every case per study protocol including dengue RDT (NS1 and IgM, Standard Diagnostics, South Korea), influenza RDT (QuickVue, Quidel Corporation, USA), only for paediatric patients age < 7 years old and leptospirosis RDT (Leptospira IgM/IgG, Standard Diagnostics), only for paediatric patients age ≥ 7 years old and all adult patients

*Four multiplex real-time PCR assays detecting 15 virus subtypes of 10 viruses; Influenza (A & B), Adenovirus, Enterovirus, Respiratory syncytial virus (A & B), Metapneumovirus, Rhinovirus, Parainfluenza virus (1, 2, 3 & 4), Coronavirus, Bocavirus (subtype OC43 & NL63), and Parechovirus

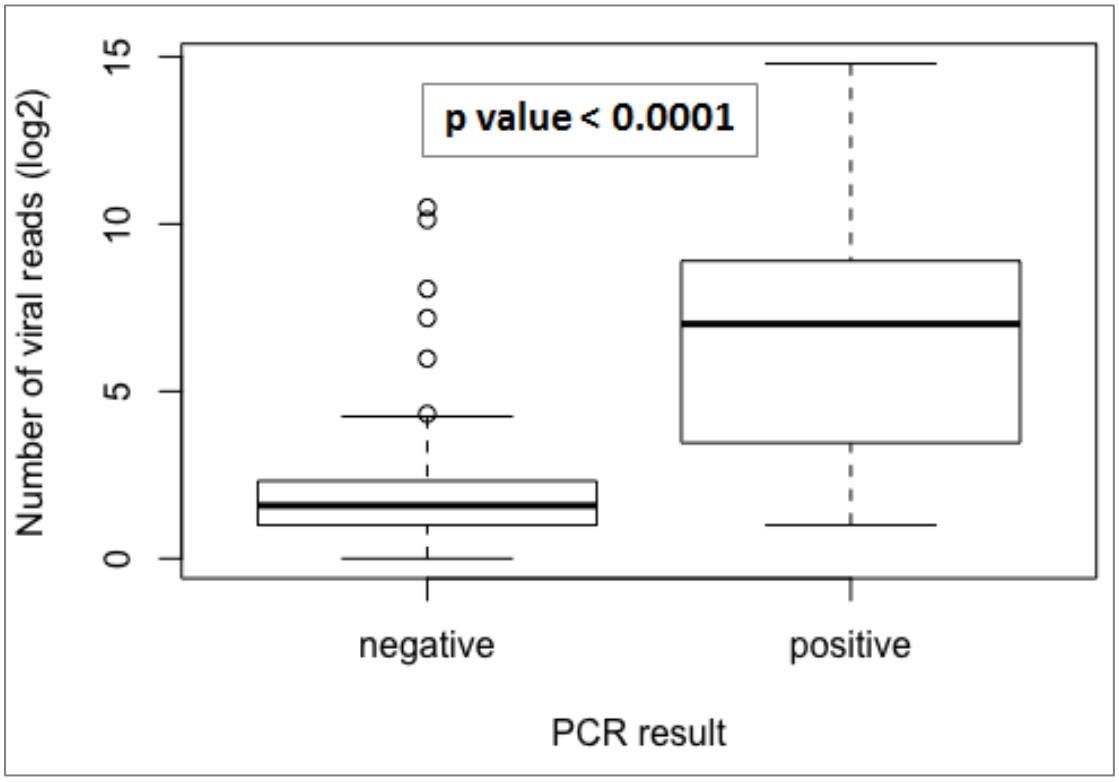
[$]5 real-time assays detecting 5 bacteria; *Legionella pneumophila*, *Mycoplasma pneumoniae*, *Chlamydophila pneumoniae*, *Chlamydophila psittaci* and *Bordetella pertussis*

Appendix 4: Bar chart showing the number of reads obtained from individual samples. Each vertical bar represents one sample. (A) Vietnam, (B) Thailand





Appendix 5: Boxplots showing the difference in the numbers of viral hits between PCR positive and negative groups



Appendix 6: The number of viral reads and genome coverage in individual samples from Vietnamese patients

Virus	Sample type	Number of total reads	Number of viral reads	E-value	Percentage of genome coverage (contig length/genomic size, bp)
Cyclovirus VN	Serum	442730	17	1.43937E-53	90.0% (1,671/1,856)
Cytomegalovirus	Swabs	493794	9	2.51E-38	>1% (624/235,403)
Cytomegalovirus	Swabs	211254	25	4.4231E-66	>1% (1,046/235,272)
Cytomegalovirus	Swabs	504418	200	1.4765E-109	1.3% (2,861/223,782)
Cytomegalovirus	Swabs	591732	12	2.90191E-52	>1% (900/235717)
Cytomegalovirus	Swabs	855702	4	5.94419E-25	>1% (309/235,272)
Cytomegalovirus	Swabs	1101990	1629	0	20.1% (44,960/223,782)
Cytomegalovirus	Swabs	1203206	53	3.20843E-51	1.9% (4,517/236,032)
Cytomegalovirus	Swabs	533616	86	8.87413E-53	2.2% (5,402/235,834)
Cytomegalovirus	Swabs	126026	204	4.7457E-111	7.1% (15,891/223,782)
Dengue virus	Serum	427286	3828	0	94.9% (10,174/10,723)
Dengue virus	Serum	560500	4721	0	95.0% (10,188/10,723)
Enterovirus	Serum	565748	405	0	42.0% (3,077/7,328)
Enterovirus	Serum	861012	6536	8.5729E-57	97.8% (6,795/6,946), 26.4% (1,941/7,345)
Enterovirus	Serum	348940	22	9.1753E-53	27.9% (1,841/6,606)
Enterovirus	Swabs	711594	329	0	31.4% (2,318/7,345)
Enterovirus	Swabs	493794	11	8.23215E-53	8% (579/7,345bp)
Enterovirus	Swabs	876786	29	1.5706E-155	61.8% (4,390/7,104)
Enterovirus	Serum	443284	170	5.827E-180	45.5% (3,379/7,434)
Enterovirus	Swabs	904408	6	2.31739E-22	2% (146/7,206)
Enterovirus	Swabs	825274	787	0	48.5% (3,204/6,612)
Enterovirus	Serum	172824	131	1.5905E-169	19.1% (1,421/7,432)
Enterovirus	Serum	200880	14	7.2996E-36	7.2% (537/7,427)
Enterovirus	Serum	349020	166	0	13.2% (980/7,433)
Enterovirus	Serum	205366	184	0	10.7% (703/6,591)
Enterovirus	Swabs	102766	5	1.30753E-39	2% (127/7,368)
Epstein-Barr virus	Serum	755526	4	2.59398E-36	<1% (336/169,864)
Epstein-Barr virus	Swabs	604874	3	1.98845E-32	1.6% (2,789/169,864)
Epstein-Barr virus	Swabs	904408	2	4.13442E-11	1.6% (2,795/169,864)
Epstein-Barr virus	Swabs	732950	2	1.18234E-12	1.5% (2,607/169,864)
Epstein-Barr virus	Swabs	126026	6	3.45686E-24	1.7% (2,905/169,864)
Gemycircularvirus SL1	Serum	381902	1668	3.06932E-60	100% (2,199)
Gemycircularvirus SL1	Swabs	883776	23	1.4508E-112	77.2% (1,697/2,199)
Gemycircularvirus SL1	Serum	442730	41	1.49764E-60	3.8% (85/2,199)
Gemycircularvirus SL1	Serum	281200	11	1.649E-102	52.5% (1,156/2,199)
Gemycircularvirus SL1	Swabs	219956	2	1.79015E-49	22.5% (494/2,199)
Hepatitis B virus	Serum	11076	183	5.56485E-51	94.7% (3,044/3,215)
Hepatitis B virus	Stool	441248	127	1.074E-142	78.7% (2,529/3,215)
Hepatitis B virus	Serum	560500	2	2.66806E-54	9.3% (299/3,215)
Hepatitis B virus	Serum	649082	2	1.08203E-30	5% (175/3,215)
Hepatitis B virus	Serum	298130	22918	7.56657E-56	100% (3,215)
Hepatitis B virus	Serum	352212	982	6.7019E-158	95.6% (3,074/3,215)
Hepatitis B virus	Serum	438146	18364	3.42494E-56	100% (3,215)
Hepatitis B virus	Serum	54750	1732	5.6804E-124	100% (3,215)
Hepatitis B virus	Serum	374364	477	6.3359E-179	90.6% (2,914/3,215)

Hepatitis B virus	Serum	320784	75	0	84% (2,709/3,215)
Hepatitis C virus	Serum	293316	5342	1.40256E-56	98.6% (9,171/9,297)
Hepatitis C virus	Serum	231960	307	2.06237E-50	74.6% (6,984/9,358)
Human coronavirus	Swabs	1203206	4	1.86221E-28	1.1% (328/30,521)
Human herpesvirus 6	Swabs	1050246	16	8.72727E-53	0.8% (1,309/161,296)
Human immunodeficiency virus	Serum	293316	355	8.4618E-102	55.1% (4,883/8,860)
Human mastadenovirus	Swabs	855702	6	5.02331E-30	1.6% (582/35,831)
Human mastadenovirus	Swabs	975450	287	0	36.2% (12,774/35,265)
Human metapneumovirus	Swabs	825274	522	0	74.5% (9,932/13,327)
Human parainfluenza virus	Swabs	701436	3	8.16848E-51	2.5% (393/15,502)
Human parainfluenza virus	Swabs	111152	427	0	59.0% (9,047/15,335)
Human parechovirus	Serum	331722	58	0	15.8% (1,155/7,320)
Human parechovirus	Stool	608352	52	7.29444E-96	16.2% (1,186/7,320)
Human pegivirus 2	Serum	293316	273	0	33.9% (3,237/9,538)
Human respiratory syncytial virus	Swabs	504418	28422	0	99.3% (15,165/15,276)
Human respiratory syncytial virus	Swabs	452112	9	3.04048E-35	3% (467/15,232)
Human rhinovirus	Serum	513280	483	0	59.4% (4,217/7,099)
Human rhinovirus	Swabs	811032	401	0	14.6% (974/6,692bp)
Human rhinovirus	Swabs	408734	8	3.66242E-55	5.4% (387/7,208bp)
Human rhinovirus	Swabs	489110	39	5.99188E-55	25.0% (1,761/7,047)
Human rhinovirus	Swabs	876786	67	0	61.8% (4,390/7,104)
Influenza A virus	Swabs	493794	23	3.39944E-53	5.2% (710/13,500)
Influenza B virus	Swabs	479434	594	6.7272E-145	58.4% (1,025/1,755)
Measles virus	Stool	441248	19530	0	96.6% (15,360/15,894)
Measles virus	Stool	435356	4	2.60588E-49	3.8% (602/15,894)
Rotavirus A	Serum	960504	2	5.84582E-25	<1% (155/18,550)
Rotavirus A	Serum	590870	2	1.59659E-12	<1% (100/18,550)
Rotavirus A	Serum	491942	366	0	83.0% (2,731/3,292)
Saffold virus	Swabs	1050246	29	4.5541E-138	9.2% (737/8,054)
Salivirus A	Stool	435356	4	1.95271E-37	7.3% (582/8,021)
WU Polyomavirus	Swabs	459132	164	1.1621E-129	45.3% (2,367/5,229)

Appendix 7: List of common contaminants and viruses not reported in human samples

Viral family	Species	Genome	Number of matching reads	Best BLASTx E value	Detected in (n)	Other virus found	Patient group	Related species was previously reported in	References
Adenoviridae	Bovine mastadenovirus C	dsDNA	4	3.47E-12	Serum (1)		Adults	Cattle	(325)
Coronaviridae	Bulbμl coronavirus HKU11	ssRNA	2	2.92E-06	Serum (1)		Adults	Wild bird	(326)
Coronaviridae	Penaeus monodon circovirus VN11	ssRNA	2,5&15	9.46E-05	Pooled swabs (1) Serum (2)		Children and Adults	Numerous including pigs	(327)
Nodaviridae	Nodamura virus	ssRNA	2	9.13E-11	Stool (1)	Shuangao insect virus 11	Adults	Insects	(328)
Picornaviridae	Boone cardiovirus	ssRNA	17	5.45E-07	Serum (1)		Children	Rats	(329)
Picobirnaviridae	Dromedary picobirnavirus	dsRNA	8	4.90E-87	Pooled swabs (1)		Adults	Camels	(330)
Picobirnaviridae	Otarine_picobirnavirus	dsRNA	589&2	0	Pooled sera (1) Pooled swabs (1)		Adult	Sea lion	(331)
Parvoviridae	Bat parvovirus	ssDNA	3	1.22E-08	Serum (2)		Children	Bat	(332)
Papillomaviridae	Human papillomavirus	circular dsDNA	4	2.86217E-42	Serum (1)		Adult		
Parvoviridae	Densovirus	ssDNA	≤ 559	5.31243E-96	Pooled swabs (12) Serum (38) CSF (1)		Children and Adults	Mosquitoes	(333)
Partitiviridae	Partitivirus	dsRNA	≤ 152	2.0824e-103	Pooled swabs (4) Serum (22)		Children and Adults	Fungi	(334)
Parvoviridae	Parvovirus NIH-CQV	ssDNA	≤ 104	2.43E-52	Pooled swabs (4) Serum (87) Stool (1) CSF (3)		Children and Adults	Qiagen column contaminant	(248)
Reoviridae	Kadipiro virus	dsRNA	3	7.87E-32	Serum (5)		Adults and Children	Contaminant	(249,335)
Reoviridae	Lutzomyia reovirus 1	dsRNA	24	5.54E-07	Serum (1)		Adults	Sand flies	(336)
Reoviridae	Eubenangee virus	dsRNA	1	9.23E-05	Serum (1)	Tilligerry virus	Adults	Marsupials, cattle, mosquitoes and *Culicoides*	(337)
Reoviridae	Cypovirus	dsRNA	1,4,7	4.03196E-40	Serum (3)		Children and Adults	Insect	(338)
Rhabdoviridae	Curionopolis virus	ssRNA	6	4.45E-05	Serum (1)		Children	Culicoides	(339)
Siphoviridae	Streptococcus_virus_MS1	DNA	12&22	7.2616E-54	Pooled swabs (1) Pooled stool (1)				(340)
Totiviridae	Saccharomyces cerevisiae virus L-BC (La)	dsRNA	9	7.45478E-21	Pooled swabs (1) Serum (3)		Children and Adults	Fungi	(341)
Totiviridae	Saccharomyces_cerevisiae_virus_L-A	dsRNA	8	1.51379E-23	Pooled sera (1)		Children and Adults	Yeast	

Totiviridae	Scheffersomyces segobiensis virus L	dsRNA	≤ 857	3.50723E-118	Pooled swabs (5) Serum (22) Stool (2) CSF (1)		Children and Adults	Fungi	(342)
Unclassified	Magnaporthe oryzae RNA virus	RNA	9,14,15& 21	1.06135E-39	Serum (4)		Children and Adults	Fungi	(343)
Unclassified	Mosquito VEM virus SDRBAJ	ssDNA	3	1.40E-07	Serum (1)	CRESS virus	Adults	red snapper tissue	Unpublished paper
Unclassified	Nepavirus	ssDNA	2	6.44E-06	Serum (1)		Children	Untreated Sewage	(344)

Appendix 8: Genetic distances of different genes of HPgV-2 at nucleic acid level (%). Note: VN: Vietnam, CHN: China, U.S.: United States, UK: United of Kingdom

S	U.S.	U.S.	VN	VN	VN	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	UK	UK	U.S	CHN	CHN
U.S.		94.93	96.77	97.7	98.04	96.77	96.77	97.7	97.7	94.47	94.47	94.47	96.31	97.7	97.7	97.7	97.93	97.24	97.7	97.7	97.7	97.7	96.54	96.31	96.31	96.54	100	94.93
U.S.	94.93		95.39	96.31	96.08	94.47	94.47	96.31	96.31	94.01	94.01	94.01	96.77	96.31	95.39	95.39	96.54	96.77	96.31	96.31	96.31	96.31	96.54	95.85	94.93	97	94.93	100
VN	96.77	95.39		98.16	98.04	97.24	97.24	98.16	98.16	95.85	95.85	95.85	97.7	99.08	96.31	96.31	98.39	98.62	99.08	99.08	99.08	99.08	97	96.77	97.7	97	96.77	95.39
VN	97.7	96.31	98.16		98.04	97.24	97.24	98.16	98.16	95.85	95.85	95.85	97.7	99.08	97.7	97.7	98.39	97.7	99.08	99.08	99.08	99.08	96.54	96.77	97.7	97	97.7	96.31
VN	98.04	96.08	98.04	98.04		96.08	96.08	99.02	99.02	93.14	93.14	93.14	96.08	99.02	98.04	98.04	99.51	100	99.02	99.02	99.02	99.02	99.51	98.04	96.08	96.57	98.04	96.08
U.S.	96.77	94.47	97.24	97.24	96.08		100	97.24	97.24	94.93	94.93	94.93	96.77	98.16	95.39	95.39	97.47	96.77	98.16	98.16	98.16	98.16	96.08	95.85	96.77	96.08	96.77	94.47
U.S.	96.77	94.47	97.24	97.24	96.08	100		97.24	97.24	94.93	94.93	94.93	96.77	98.16	95.39	95.39	97.47	96.77	98.16	98.16	98.16	98.16	96.08	95.85	96.77	96.08	96.77	94.47
U.S.	97.7	96.31	98.16	98.16	99.02	97.24	97.24		100	95.85	95.85	95.85	97.7	99.08	97.24	97.24	99.31	98.62	99.08	99.08	99.08	99.08	97.47	97.7	97.7	97.93	97.7	96.31
U.S.	97.7	96.31	98.16	98.16	99.02	97.24	97.24	100		95.85	95.85	95.85	97.7	99.08	97.24	97.24	99.31	98.62	99.08	99.08	99.08	99.08	97.47	97.7	97.7	97.93	97.7	96.31
U.S.	94.47	94.01	95.85	95.85	93.14	94.93	94.93	95.85	95.85		100	100	96.31	96.77	94.47	94.47	96.54	95.39	96.77	96.77	96.77	96.77	94.7	95.39	97.24	97	94.47	94.01
U.S.	94.47	94.01	95.85	95.85	93.14	94.93	94.93	95.85	95.85	100		100	96.31	96.77	94.47	94.47	96.54	95.39	96.77	96.77	96.77	96.77	94.7	95.39	97.24	97	94.47	94.01
U.S.	94.47	94.01	95.85	95.85	93.14	94.93	94.93	95.85	95.85	100	100		96.31	96.77	94.47	94.47	96.54	95.39	96.77	96.77	96.77	96.77	94.7	95.39	97.24	97	94.47	94.01
U.S.	96.31	96.77	97.7	97.7	96.08	96.77	96.77	97.7	97.7	96.31	96.31	96.31		98.62	95.85	95.85	97.93	97.24	98.62	98.62	98.62	98.62	96.08	96.31	97.24	98.39	96.31	96.77
U.S.	97.7	96.31	99.08	99.08	99.02	98.16	98.16	99.08	99.08	96.77	96.77	96.77	98.62		97.24	97.24	99.31	98.62	100	100	100	100	97.47	97.7	98.62	97.93	97.7	96.31
U.S.	97.7	95.39	96.31	97.7	98.04	95.39	95.39	97.24	97.24	94.47	94.47	94.47	95.85	97.24		100	97.47	96.77	97.24	97.24	97.24	97.24	96.08	96.31	96.31	96.08	97.7	95.39
U.S.	97.7	95.39	96.31	97.7	98.04	95.39	95.39	97.24	97.24	94.47	94.47	94.47	95.85	97.24	100		97.47	96.77	97.24	97.24	97.24	97.24	96.08	96.31	96.31	96.08	97.7	95.39
U.S.	97.93	96.54	98.39	98.39	99.51	97.47	97.47	99.31	99.31	96.54	96.54	96.54	97.93	99.31	97.47	97.47		98.85	99.31	99.31	99.31	99.31	97.93	97.93	97.93	98.39	97.93	96.54
U.S.	97.24	96.77	98.62	97.7	100	96.77	96.77	98.62	98.62	95.39	95.39	95.39	97.24	98.62	96.77	96.77	98.85		98.62	98.62	98.62	98.62	98.39	97.24	97.24	97.47	97.24	96.77
U.S.	97.7	96.31	99.08	99.08	99.02	98.16	98.16	99.08	99.08	96.77	96.77	96.77	98.62	100	97.24	97.24	99.31	98.62		100	100	100	97.47	97.7	98.62	97.93	97.7	96.31
U.S.	97.7	96.31	99.08	99.08	99.02	98.16	98.16	99.08	99.08	96.77	96.77	96.77	98.62	100	97.24	97.24	99.31	98.62	100		100	100	97.47	97.7	98.62	97.93	97.7	96.31
U.S.	97.7	96.31	99.08	99.08	99.02	98.16	98.16	99.08	99.08	96.77	96.77	96.77	98.62	100	97.24	97.24	99.31	98.62	100	100		100	97.47	97.7	98.62	97.93	97.7	96.31
U.S.	97.7	96.31	99.08	99.08	99.02	98.16	98.16	99.08	99.08	96.77	96.77	96.77	98.62	100	97.24	97.24	99.31	98.62	100	100	100		97.47	97.7	98.62	97.93	97.7	96.31
U.S.	96.54	96.54	97	96.54	99.51	96.08	96.08	97.47	97.47	94.7	94.7	94.7	96.08	97.47	96.08	96.08	97.93	98.39	97.47	97.47	97.47	97.47		97	96.08	96.54	96.54	96.54
UK	96.31	95.85	96.77	96.77	98.04	95.85	95.85	97.7	97.7	95.39	95.39	95.39	96.31	97.7	96.31	96.31	97.93	97.24	97.7	97.7	97.7	97.7	97		96.31	96.54	96.31	95.85
UK	96.31	94.93	97.7	97.7	96.08	96.77	96.77	97.7	97.7	97.24	97.24	97.24	97.24	98.62	96.31	96.31	97.93	97.24	98.62	98.62	98.62	98.62	96.08	96.31		97.47	96.31	94.93
U.S.	96.54	97	97	97	96.57	96.08	96.08	97.93	97.93	97	97	97	98.39	97.93	96.08	96.08	98.39	97.47	97.93	97.93	97.93	97.93	96.54	96.54	97.47		96.54	97
CHN	100	94.93	96.77	97.7	98.04	96.77	96.77	97.7	97.7	94.47	94.47	94.47	96.31	97.7	97.7	97.7	97.93	97.24	97.7	97.7	97.7	97.7	96.54	96.31	96.31	96.54		94.93
CHN	94.93	100	95.39	96.31	96.08	94.47	94.47	96.31	96.31	94.01	94.01	94.01	96.77	96.31	95.39	95.39	96.54	96.77	96.31	96.31	96.31	96.31	96.54	95.85	94.93	97	94.93	

E1	VN	VN	VN	VN	VN	**U.S**	**U.S**	**U.S**	**U.S**	**U.S**	**U.S**	**U.S**	**U.S**	**U.S**	**U.S**	**U.S**	**U.S**	**U.S**	**U.S**	**U.S**	**U.S**	**U.S**	**U.S**	UK	UK	**U.S**	CHN	CHN
VN		95.99	96.68	94.24	93.02	95.11	95.11	95.29	95.29	95.29	95.29	95.29	96.34	97.21	94.85	94.85	95.11	95.81	95.29	95.29	95.29	95.29	95.29	93.02	94.94	95.64	92.32	93.72
VN	95.99		96.51	94.04	92.18	94.24	94.24	94.24	94.24	93.54	93.54	93.54	95.81	95.99	93.98	93.98	94.42	94.59	94.59	94.59	94.59	94.59	93.54	91.97	94.07	95.29	92.67	92.67
VN	96.68	96.51		94.44	92.39	94.59	94.59	94.94	94.94	93.89	93.89	93.89	96.16	96.34	93.8	93.8	94.59	94.24	95.64	95.64	95.64	95.64	93.89	91.97	94.24	95.29	93.72	92.84
VN	94.24	94.04	94.44		93.34	94.24	94.24	95.45	95.45	95.45	95.45	95.45	96.46	96.67	94.14	94.14	94.04	94.44	94.85	94.85	94.85	94.85	93.84	92.42	93.84	95.86	92.42	93.64
VN	93.02	92.18	92.39	93.34		94.08	94.08	94.29	94.29	93.45	93.45	93.45	94.93	95.14	92.07	92.07	93.23	95.14	93.02	93.02	93.02	93.02	92.81	91.54	91.97	94.29	91.12	93.23
U.S.	95.11	94.24	94.59	94.24	94.08		100	95.29	95.29	95.81	95.81	95.81	96.34	97.21	94.68	94.68	94.76	95.29	95.29	95.29	95.29	95.29	93.72	93.72	95.11	96.16	93.72	93.89
U.S.	95.11	94.24	94.59	94.24	94.08	100		95.29	95.29	95.81	95.81	95.81	96.34	97.21	94.68	94.68	94.76	95.29	95.29	95.29	95.29	95.29	93.72	93.72	95.11	96.16	93.72	93.89
U.S.	95.29	94.24	94.94	95.45	94.29	95.29	95.29		100	95.29	95.29	95.29	97.03	97.38	95.2	95.2	95.81	95.64	95.46	95.46	95.46	95.46	95.29	93.54	94.24	96.68	94.07	94.59
U.S.	95.29	94.24	94.94	95.45	94.29	95.29	95.29	100		95.29	95.29	95.29	97.03	97.38	95.2	95.2	95.81	95.64	95.46	95.46	95.46	95.46	95.29	93.54	94.24	96.68	94.07	94.59
U.S.	95.29	93.54	93.89	95.45	93.45	95.81	95.81	95.29	95.29		100	100	96.68	97.56	94.33	94.33	94.59	94.76	94.76	94.76	94.76	94.76	95.46	93.89	95.11	96.34	93.02	93.02
U.S.	95.29	93.54	93.89	95.45	93.45	95.81	95.81	95.29	95.29	100		100	96.68	97.56	94.33	94.33	94.59	94.76	94.76	94.76	94.76	94.76	95.46	93.89	95.11	96.34	93.02	93.02
U.S.	95.29	93.54	93.89	95.45	93.45	95.81	95.81	95.29	95.29	100	100		96.68	97.56	94.33	94.33	94.59	94.76	94.76	94.76	94.76	94.76	95.46	93.89	95.11	96.34	93.02	93.02
U.S.	96.34	95.81	96.16	96.46	94.93	96.34	96.34	97.03	97.03	96.68	96.68	96.68		98.78	95.2	95.2	96.16	96.68	97.56	97.56	97.56	97.56	96.34	94.24	95.99	97.73	94.94	95.11
U.S.	97.21	95.99	96.34	96.67	95.14	97.21	97.21	97.38	97.38	97.56	97.56	97.56	98.78		96.42	96.42	96.16	97.21	97.21	97.21	97.21	97.21	96.51	95.11	96.16	98.08	94.76	95.11
U.S.	94.85	93.98	93.8	94.14	92.07	94.68	94.68	95.2	95.2	94.33	94.33	94.33	95.2	96.42		99.91	93.28	94.33	94.5	94.5	94.5	94.5	94.33	92.58	95.03	95.55	92.76	93.63
U.S.	94.85	93.98	93.8	94.14	92.07	94.68	94.68	95.2	95.2	94.33	94.33	94.33	95.2	96.42	99.91		93.28	94.33	94.5	94.5	94.5	94.5	94.33	92.58	95.03	95.55	92.76	93.63
U.S.	95.11	94.42	94.59	94.04	93.23	94.76	94.76	95.81	95.81	94.59	94.59	94.59	96.16	96.16	93.28	93.28		94.76	95.29	95.29	95.29	95.29	93.89	92.84	93.89	96.16	93.19	93.02
U.S.	95.81	94.59	94.24	94.44	95.14	95.29	95.29	95.64	95.64	94.76	94.76	94.76	96.68	97.21	94.33	94.33	94.76		95.11	95.11	95.11	95.11	95.11	93.89	94.07	95.99	93.37	94.76
U.S.	95.29	94.59	95.64	94.85	93.02	95.29	95.29	95.46	95.46	94.76	94.76	94.76	97.56	97.21	94.5	94.5	95.29	95.11		100	100	100	94.24	93.37	94.76	96.86	94.42	93.37
U.S.	95.29	94.59	95.64	94.85	93.02	95.29	95.29	95.46	95.46	94.76	94.76	94.76	97.56	97.21	94.5	94.5	95.29	95.11	100		100	100	94.24	93.37	94.76	96.86	94.42	93.37
U.S.	95.29	94.59	95.64	94.85	93.02	95.29	95.29	95.46	95.46	94.76	94.76	94.76	97.56	97.21	94.5	94.5	95.29	95.11	100	100		100	94.24	93.37	94.76	96.86	94.42	93.37
U.S.	95.29	94.59	95.64	94.85	93.02	95.29	95.29	95.46	95.46	94.76	94.76	94.76	97.56	97.21	94.5	94.5	95.29	95.11	100	100	100		94.24	93.37	94.76	96.86	94.42	93.37
U.S.	95.29	93.54	93.89	93.84	92.81	93.72	93.72	95.29	95.29	95.46	95.46	95.46	96.34	96.51	94.33	94.33	93.89	95.11	94.24	94.24	94.24	94.24		93.02	95.11	94.94	91.97	93.89
UK	93.02	91.97	91.97	92.42	91.54	93.72	93.72	93.54	93.54	93.89	93.89	93.89	94.24	95.11	92.58	92.58	92.84	93.89	93.37	93.37	93.37	93.37	93.02		92.84	93.72	92.32	93.02
UK	94.94	94.07	94.24	93.84	91.97	95.11	95.11	94.24	94.24	95.11	95.11	95.11	95.99	96.16	95.03	95.03	93.89	94.07	94.76	94.76	94.76	94.76	95.11	92.84		95.29	93.54	93.89
U.S.	95.64	95.29	95.29	95.86	94.29	96.16	96.16	96.68	96.68	96.34	96.34	96.34	97.73	98.08	95.55	95.55	96.16	95.99	96.86	96.86	96.86	96.86	94.94	93.72	95.29		94.42	94.76
CHN	92.32	92.67	93.72	92.42	91.12	93.72	93.72	94.07	94.07	93.02	93.02	93.02	94.94	94.76	92.76	92.76	93.19	93.37	94.42	94.42	94.42	94.42	91.97	92.32	93.54	94.42		91.97
CHN	93.72	92.67	92.84	93.64	93.23	93.89	93.89	94.59	94.59	93.02	93.02	93.02	95.11	95.11	93.63	93.63	93.02	94.76	93.37	93.37	93.37	93.37	93.89	93.02	93.89	94.76	91.97	

E2	VN	VN	VN	VN	VN	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	UK	UK	U.S	CHN	CHN
VN		95.75	95.9	93.77	94.2	93.4	93.4	93.77	93.77	92.55	92.55	92.55	95.38	94.36	93.07	93.07	92.08	93.58	94.53	94.53	94.53	94.53	91.79	92.36	93.09	92.92	91.98	94.72
VN	95.75		97.31	94.53	95.14	93.58	93.58	94.72	94.72	93.21	93.21	93.21	95.75	94.46	93.25	93.25	92.26	94.06	94.43	94.43	94.43	94.43	92.55	92.92	93	93.02	93.02	95.57
VN	95.9	97.31		94.67	95.19	94.58	94.58	94.48	94.48	93.35	93.35	93.35	96.18	95.35	94.15	94.15	92.78	94.86	95.42	95.42	95.42	95.42	93.21	93.35	93.7	93.82	92.78	95.33
VN	93.77	94.53	94.67		94.67	93.21	93.21	94.34	94.34	92.64	92.64	92.64	95.75	94.17	93.73	93.73	91.89	94.06	93.77	93.77	93.77	93.77	92.74	93.02	92.81	93.3	92.83	94.43
VN	94.2	95.14	95.19	94.67		93.96	93.96	94.95	94.95	93.63	93.63	93.63	96.18	94.88	95	95	92.59	94.48	94.76	94.76	94.76	94.76	93.63	93.44	93.89	93.92	93.16	95.14
U.S.	93.4	93.58	94.58	93.21	93.96		99.15	93.77	93.77	93.11	93.11	93.11	95.28	98.04	93.92	93.92	91.89	93.49	97.83	97.83	97.83	97.83	93.11	92.83	92.38	93.35	92.26	93.73
U.S.	93.4	93.58	94.58	93.21	93.96	99.15		93.77	93.77	93.11	93.11	93.11	95.28	98.04	93.92	93.92	91.89	93.49	97.83	97.83	97.83	97.83	93.11	92.83	92.38	93.35	92.26	93.73
U.S.	93.77	94.72	94.48	94.34	94.95	93.77	93.77		100	93.21	93.21	93.21	96.13	94.55	94.81	94.81	92.45	94.34	93.96	93.96	93.96	93.96	93.58	93.58	93.28	94.15	93.49	95.28
U.S.	93.77	94.72	94.48	94.34	94.95	93.77	93.77	100		93.21	93.21	93.21	96.13	94.55	94.81	94.81	92.45	94.34	93.96	93.96	93.96	93.96	93.58	93.58	93.28	94.15	93.49	95.28
U.S.	92.55	93.21	93.35	92.64	93.63	93.11	93.11	93.21	93.21		100	100	94.34	94.65	92.78	92.78	92.08	92.45	93.58	93.58	93.58	93.58	92.64	92.08	92.24	92.92	92.83	92.45
U.S.	92.55	93.21	93.35	92.64	93.63	93.11	93.11	93.21	93.21	100		100	94.34	94.65	92.78	92.78	92.08	92.45	93.58	93.58	93.58	93.58	92.64	92.08	92.24	92.92	92.83	92.45
U.S.	92.55	93.21	93.35	92.64	93.63	93.11	93.11	93.21	93.21	100	100		94.34	94.65	92.78	92.78	92.08	92.45	93.58	93.58	93.58	93.58	92.64	92.08	92.24	92.92	92.83	92.45
U.S.	95.38	95.75	96.18	95.75	96.18	95.28	95.28	96.13	96.13	94.34	94.34	94.34		96.34	95.24	95.24	93.3	95.57	95.85	95.85	95.85	95.85	94.43	95.19	94.5	95.38	94.15	96.04
U.S.	94.36	94.46	95.35	94.17	94.88	98.04	98.04	94.55	94.55	94.65	94.65	94.65	96.34		95.35	95.35	92.48	94.27	98.61	98.61	98.61	98.61	93.61	93.8	93.3	94.17	93.51	94.46
U.S.	93.07	93.25	94.15	93.73	95	93.92	93.92	94.81	94.81	92.78	92.78	92.78	95.24	95.35		99.76	92.59	93.92	94.2	94.2	94.2	94.2	93.63	92.69	92.95	93.82	93.4	93.73
U.S.	93.07	93.25	94.15	93.73	95	93.92	93.92	94.81	94.81	92.78	92.78	92.78	95.24	95.35	99.76		92.59	93.92	94.2	94.2	94.2	94.2	93.63	92.69	92.95	93.82	93.4	93.73
U.S.	92.08	92.26	92.78	91.89	92.59	91.89	91.89	92.45	92.45	92.08	92.08	92.08	93.3	92.48	92.59	92.59		92.26	92.17	92.17	92.17	92.17	92.17	91.42	91.39	91.7	91.6	92.17
U.S.	93.58	94.06	94.86	94.06	94.48	93.49	93.49	94.34	94.34	92.45	92.45	92.45	95.57	94.27	93.92	93.92	92.26		94.15	94.15	94.15	94.15	93.4	92.92	93.75	94.34	92.26	95.09
U.S.	94.53	94.43	95.42	93.77	94.76	97.83	97.83	93.96	93.96	93.58	93.58	93.58	95.85	98.61	94.2	94.2	92.17	94.15		100	100	100	93.11	93.11	93.66	93.4	92.83	94.53
U.S.	94.53	94.43	95.42	93.77	94.76	97.83	97.83	93.96	93.96	93.58	93.58	93.58	95.85	98.61	94.2	94.2	92.17	94.15	100		100	100	93.11	93.11	93.66	93.4	92.83	94.53
U.S.	94.53	94.43	95.42	93.77	94.76	97.83	97.83	93.96	93.96	93.58	93.58	93.58	95.85	98.61	94.2	94.2	92.17	94.15	100	100		100	93.11	93.11	93.66	93.4	92.83	94.53
U.S.	94.53	94.43	95.42	93.77	94.76	97.83	97.83	93.96	93.96	93.58	93.58	93.58	95.85	98.61	94.2	94.2	92.17	94.15	100	100	100		93.11	93.11	93.66	93.4	92.83	94.53
U.S.	91.79	92.55	93.21	92.74	93.63	93.11	93.11	93.58	93.58	92.64	92.64	92.64	94.43	93.61	93.63	93.63	92.17	93.4	93.11	93.11	93.11	93.11		92.26	92.15	93.96	91.98	92.64
UK	92.36	92.92	93.35	93.02	93.44	92.83	92.83	93.58	93.58	92.08	92.08	92.08	95.19	93.8	92.69	92.69	91.42	92.92	93.11	93.11	93.11	93.11	92.26		92.43	92.92	92.64	93.11
UK	93.09	93	93.7	92.81	93.89	92.38	92.38	93.28	93.28	92.24	92.24	92.24	94.5	93.3	92.95	92.95	91.39	93.75	93.66	93.66	93.66	93.66	92.15	92.43		92.81	92.43	94.03
U.S.	92.92	93.02	93.82	93.3	93.92	93.35	93.35	94.15	94.15	92.92	92.92	92.92	95.38	94.17	93.82	93.82	91.7	94.34	93.4	93.4	93.4	93.4	93.96	92.92	92.81		92.64	93.58
CHN	91.98	93.02	92.78	92.83	93.16	92.26	92.26	93.49	93.49	92.83	92.83	92.83	94.15	93.51	93.4	93.4	91.6	92.26	92.83	92.83	92.83	92.83	91.98	92.64	92.43	92.64		93.11
CHN	94.72	95.57	95.33	94.43	95.14	93.73	93.73	95.28	95.28	92.45	92.45	92.45	96.04	94.46	93.73	93.73	92.17	95.09	94.53	94.53	94.53	94.53	92.64	93.11	94.03	93.58	93.11	

X	CHN	CHN	VN	VN	VN	VN	VN	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	UK	UK	U.S
CHN		89.9	89.34	89.76	90.11	91.8	90.11	91.16	91.16	91.73	91.73	92.15	92.15	92.15	92.85	92.53	89.13	89.13	90.46	90.32	91.44	91.44	91.44	91.44	91.16	92.15	91.02	91.37
CHN	89.9		91.87	91.73	91.94	92.64	91.94	91.3	91.3	91.02	91.02	92.71	92.71	92.71	92.85	93.23	90.95	90.95	90.88	91.73	92.01	92.01	92.01	92.01	91.87	92.15	91.58	92.43
VN	89.34	91.87		94.95	95.3	90.39	91.09	90.18	90.18	91.58	91.58	90.32	90.32	90.32	92.15	92.53	90.32	90.32	89.76	91.3	92.01	92.01	92.01	92.01	90.18	90.74	90.74	91.23
VN	89.76	91.73	94.95		95.16	91.09	90.81	90.88	90.88	91.3	91.3	91.44	91.44	91.44	92.71	92.95	89.76	89.76	89.9	91.44	91.58	91.58	91.58	91.58	90.88	91.44	91.16	90.81
VN	90.11	91.94	95.3	95.16		91.73	91.58	91.23	91.23	92.57	92.57	91.8	91.8	91.8	93.2	93.58	90.53	90.53	89.97	92.08	92.08	92.08	92.08	92.08	91.23	91.51	91.65	92.01
VN	91.8	92.64	90.39	91.09	91.73		92.01	93.34	93.34	93.2	93.2	93.9	93.9	93.9	94.88	95.13	90.46	90.46	93.2	92.64	93.34	93.34	93.34	93.34	93.34	93.62	93.06	93.13
VN	90.11	91.94	91.09	90.81	91.58	92.01		91.8	91.8	91.51	91.51	91.51	91.51	91.51	93.2	93.02	89.41	89.41	91.23	92.22	91.8	91.8	91.8	91.8	91.09	91.51	92.22	92.01
U.S.	91.16	91.3	90.18	90.88	91.23	93.34	91.8		99.58	93.27	93.27	92.71	92.71	92.71	94.11	95.2	90.46	90.46	91.16	91.23	92.57	92.57	92.57	92.57	91.3	91.87	91.87	92.08
U.S.	91.16	91.3	90.18	90.88	91.23	93.34	91.8	99.58		93.27	93.27	92.71	92.71	92.71	94.11	95.2	90.46	90.46	91.16	91.23	92.57	92.57	92.57	92.57	91.3	91.87	91.87	92.08
U.S.	91.73	91.02	91.58	91.3	92.57	93.2	91.51	93.27	93.27		100	93.83	93.83	93.83	94.39	94.78	90.46	90.46	92.85	92.29	93.55	93.55	93.55	93.55	92.85	92.71	92.71	92.36
U.S.	91.73	91.02	91.58	91.3	92.57	93.2	91.51	93.27	93.27	100		93.83	93.83	93.83	94.39	94.78	90.46	90.46	92.85	92.29	93.55	93.55	93.55	93.55	92.85	92.71	92.71	92.36
U.S.	92.15	92.71	90.32	91.44	91.8	93.9	91.51	92.71	92.71	93.83	93.83		100	100	96.49	96.74	92.29	92.29	93.13	93.13	94.53	94.53	94.53	94.53	93.69	94.39	93.83	93.62
U.S.	92.15	92.71	90.32	91.44	91.8	93.9	91.51	92.71	92.71	93.83	93.83	100		100	96.49	96.74	92.29	92.29	93.13	93.13	94.53	94.53	94.53	94.53	93.69	94.39	93.83	93.62
U.S.	92.15	92.71	90.32	91.44	91.8	93.9	91.51	92.71	92.71	93.83	93.83	100	100		96.49	96.74	92.29	92.29	93.13	93.13	94.53	94.53	94.53	94.53	93.69	94.39	93.83	93.62
U.S.	92.85	92.85	92.15	92.71	93.2	94.88	93.2	94.11	94.11	94.39	94.39	96.49	96.49	96.49		97.72	92.71	92.71	94.39	95.65	95.93	95.93	95.93	95.93	95.37	95.79	95.65	94.74
U.S.	92.53	93.23	92.53	92.95	93.58	95.13	93.02	95.2	95.2	94.78	94.78	96.74	96.74	96.74	97.72		94.57	94.57	94.07	94.64	96.74	96.74	96.74	96.74	94.78	95.2	95.2	94.42
U.S.	89.13	90.95	90.32	89.76	90.53	90.46	89.41	90.46	90.46	90.46	90.46	92.29	92.29	92.29	92.71	94.57		99.79	90.32	91.73	92.22	92.22	92.22	92.22	90.18	90.74	91.87	90.25
U.S.	89.13	90.95	90.32	89.76	90.53	90.46	89.41	90.46	90.46	90.46	90.46	92.29	92.29	92.29	92.71	94.57	99.79		90.32	91.73	92.22	92.22	92.22	92.22	90.18	90.74	91.87	90.25
U.S.	90.46	90.88	89.76	89.9	89.97	93.2	91.23	91.16	91.16	92.85	92.85	93.13	93.13	93.13	94.39	94.07	90.32	90.32		92.57	93.27	93.27	93.27	93.27	93.27	93.55	92.85	93.13
U.S.	90.32	91.73	91.3	91.44	92.08	92.64	92.22	91.23	91.23	92.29	92.29	93.13	93.13	93.13	95.65	94.64	91.73	91.73	92.57		93.13	93.13	93.13	93.13	93.13	93.55	92.99	93.06
U.S.	91.44	92.01	92.01	91.58	92.08	93.34	91.8	92.57	92.57	93.55	93.55	94.53	94.53	94.53	95.93	96.74	92.22	92.22	93.27	93.13		100	100	100	93.83	94.11	93.69	93.34
U.S.	91.44	92.01	92.01	91.58	92.08	93.34	91.8	92.57	92.57	93.55	93.55	94.53	94.53	94.53	95.93	96.74	92.22	92.22	93.27	93.13	100		100	100	93.83	94.11	93.69	93.34
U.S.	91.44	92.01	92.01	91.58	92.08	93.34	91.8	92.57	92.57	93.55	93.55	94.53	94.53	94.53	95.93	96.74	92.22	92.22	93.27	93.13	100	100		100	93.83	94.11	93.69	93.34
U.S.	91.44	92.01	92.01	91.58	92.08	93.34	91.8	92.57	92.57	93.55	93.55	94.53	94.53	94.53	95.93	96.74	92.22	92.22	93.27	93.13	100	100	100		93.83	94.11	93.69	93.34
U.S.	91.16	91.87	90.18	90.88	91.23	93.34	91.09	91.3	91.3	92.85	92.85	93.69	93.69	93.69	95.37	94.78	90.18	90.18	93.27	93.13	93.83	93.83	93.83	93.83		94.11	93.69	92.36
UK	92.15	92.15	90.74	91.44	91.51	93.62	91.51	91.87	91.87	92.71	92.71	94.39	94.39	94.39	95.79	95.2	90.74	90.74	93.55	93.55	94.11	94.11	94.11	94.11	94.11		93.55	93.27
UK	91.02	91.58	90.74	91.16	91.65	93.06	92.22	91.87	91.87	92.71	92.71	93.83	93.83	93.83	95.65	95.2	91.87	91.87	92.85	92.99	93.69	93.69	93.69	93.69	93.69	93.55		93.2
U.S.	91.37	92.43	91.23	90.81	92.01	93.13	92.01	92.08	92.08	92.36	92.36	93.62	93.62	93.62	94.74	94.42	90.25	90.25	93.13	93.06	93.34	93.34	93.34	93.34	92.36	93.27	93.2	

NS2	CHN	CHN	VN	VN	VN	VN	VN	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	UK	UK	U.S
CHN		92.64	92.92	93.47	92.78	93.54	91.67	92.71	92.71	94.17	94.17	93.19	93.19	93.19	93.89	94.17	93.61	93.61	93.06	94.17	92.64	92.64	92.64	92.64	93.19	93.33	93.89	92.64
CHN	92.64		95.14	95.56	94.31	94.79	94.03	94.24	94.24	95.56	95.56	95.97	95.97	95.97	96.25	96.67	95.21	95.21	95.14	96.53	94.86	94.86	94.86	94.86	94.72	95.69	95.56	94.17
VN	92.92	95.14		97.36	96.67	94.93	94.03	93.96	93.96	96.53	96.53	95.28	95.28	95.28	96.53	96.25	95.49	95.49	95.56	96.39	94.86	94.86	94.86	94.86	95.14	95.56	95.83	94.31
VN	93.47	95.56	97.36		97.08	95.07	94.17	95.07	95.07	96.39	96.39	95.42	95.42	95.42	96.53	96.39	95.9	95.9	95.42	96.25	95	95	95	95	95.56	95.69	95.42	94.58
VN	92.78	94.31	96.67	97.08		95.21	92.78	93.4	93.4	95.56	95.56	94.31	94.31	94.31	95.28	95.28	94.79	94.79	94.44	95.56	94.17	94.17	94.17	94.17	94.72	94.72	95	93.75
VN	93.54	94.79	94.93	95.07	95.21		94.1	94.44	94.44	96.46	96.46	95.76	95.76	95.76	96.32	96.18	95	95	95.49	96.6	95.07	95.07	95.07	95.07	95.35	95.35	96.53	94.65
VN	91.67	94.03	94.03	94.17	92.78	94.1		93.96	93.96	94.17	94.17	94.03	94.03	94.03	95.28	95.28	94.1	94.1	94.31	95.28	93.89	93.89	93.89	93.89	94.17	94.03	94.03	93.19
U.S.	92.71	94.24	93.96	95.07	93.4	94.44	93.96		99.93	95.35	95.35	94.79	94.79	94.79	95.9	96.6	95.28	95.28	95.35	95.35	94.38	94.38	94.38	94.38	94.65	94.79	95.35	93.54
U.S.	92.71	94.24	93.96	95.07	93.4	94.44	93.96	99.93		95.35	95.35	94.79	94.79	94.79	95.9	96.6	95.28	95.28	95.35	95.35	94.38	94.38	94.38	94.38	94.65	94.79	95.35	93.54
U.S.	94.17	95.56	96.53	96.39	95.56	96.46	94.17	95.35	95.35		100	96.81	96.81	96.81	97.64	97.64	96.88	96.88	96.53	96.81	96.67	96.67	96.67	96.67	96.39	96.81	97.36	95.56
U.S.	94.17	95.56	96.53	96.39	95.56	96.46	94.17	95.35	95.35	100		96.81	96.81	96.81	97.64	97.64	96.88	96.88	96.53	96.81	96.67	96.67	96.67	96.67	96.39	96.81	97.36	95.56
U.S.	93.19	95.97	95.28	95.42	94.31	95.76	94.03	94.79	94.79	96.81	96.81		100	100	97.22	97.78	96.32	96.32	96.39	96.67	95.97	95.97	95.97	95.97	95.42	96.11	95.97	95.28
U.S.	93.19	95.97	95.28	95.42	94.31	95.76	94.03	94.79	94.79	96.81	96.81	100		100	97.22	97.78	96.32	96.32	96.39	96.67	95.97	95.97	95.97	95.97	95.42	96.11	95.97	95.28
U.S.	93.19	95.97	95.28	95.42	94.31	95.76	94.03	94.79	94.79	96.81	96.81	100	100		97.22	97.78	96.32	96.32	96.39	96.67	95.97	95.97	95.97	95.97	95.42	96.11	95.97	95.28
U.S.	93.89	96.25	96.53	96.53	95.28	96.32	95.28	95.9	95.9	97.64	97.64	97.22	97.22	97.22		98.19	96.88	96.88	96.94	97.78	96.81	96.81	96.81	96.81	97.08	97.5	97.22	95.56
U.S.	94.17	96.67	96.25	96.39	95.28	96.18	95.28	96.6	96.6	97.64	97.64	97.78	97.78	97.78	98.19		97.92	97.92	97.08	97.64	97.22	97.22	97.22	97.22	96.81	97.08	97.08	95.28
U.S.	93.61	95.21	95.49	95.9	94.79	95	94.1	95.28	95.28	96.88	96.88	96.32	96.32	96.32	96.88	97.92		99.79	95.9	96.46	96.04	96.04	96.04	96.04	95.76	96.04	96.04	94.51
U.S.	93.61	95.21	95.49	95.9	94.79	95	94.1	95.28	95.28	96.88	96.88	96.32	96.32	96.32	96.88	97.92	99.79		95.9	96.46	96.04	96.04	96.04	96.04	95.76	96.04	96.04	94.51
U.S.	93.06	95.14	95.56	95.42	94.44	95.49	94.31	95.35	95.35	96.53	96.53	96.39	96.39	96.39	96.94	97.08	95.9	95.9		96.67	95.28	95.28	95.28	95.28	95.69	95.56	96.39	95.14
U.S.	94.17	96.53	96.39	96.25	95.56	96.6	95.28	95.35	95.35	96.81	96.81	96.67	96.67	96.67	97.78	97.64	96.46	96.46	96.67		96.11	96.11	96.11	96.11	96.81	96.67	97.22	95.42
U.S.	92.64	94.86	94.86	95	94.17	95.07	93.89	94.38	94.38	96.67	96.67	95.97	95.97	95.97	96.81	97.22	96.04	96.04	95.28	96.11		100	100	100	96.11	96.25	95.69	93.89
U.S.	92.64	94.86	94.86	95	94.17	95.07	93.89	94.38	94.38	96.67	96.67	95.97	95.97	95.97	96.81	97.22	96.04	96.04	95.28	96.11	100		100	100	96.11	96.25	95.69	93.89
U.S.	92.64	94.86	94.86	95	94.17	95.07	93.89	94.38	94.38	96.67	96.67	95.97	95.97	95.97	96.81	97.22	96.04	96.04	95.28	96.11	100	100		100	96.11	96.25	95.69	93.89
U.S.	92.64	94.86	94.86	95	94.17	95.07	93.89	94.38	94.38	96.67	96.67	95.97	95.97	95.97	96.81	97.22	96.04	96.04	95.28	96.11	100	100	100		96.11	96.25	95.69	93.89
U.S.	93.19	94.72	95.14	95.56	94.72	95.35	94.17	94.65	94.65	96.39	96.39	95.42	95.42	95.42	97.08	96.81	95.76	95.76	95.69	96.81	96.11	96.11	96.11	96.11		95.97	95.97	94.17
UK	93.33	95.69	95.56	95.69	94.72	95.35	94.03	94.79	94.79	96.81	96.81	96.11	96.11	96.11	97.5	97.08	96.04	96.04	95.56	96.67	96.25	96.25	96.25	96.25	95.97		96.81	94.58
UK	93.89	95.56	95.83	95.42	95	96.53	94.03	95.35	95.35	97.36	97.36	95.97	95.97	95.97	97.22	97.08	96.04	96.04	96.39	97.22	95.69	95.69	95.69	95.69	95.97	96.81		95.42
U.S.	92.64	94.17	94.31	94.58	93.75	94.65	93.19	93.54	93.54	95.56	95.56	95.28	95.28	95.28	95.56	95.28	94.51	94.51	95.14	95.42	93.89	93.89	93.89	93.89	94.17	94.58	95.42	

NS3	CHN	CHN	VN	VN	VN	VN	VN	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	UK	UK	U.S
CHN		92.39	93.35	93.14	93.64	93.35	93.32	93.32	93.32	93.72	93.72	92.87	92.87	92.87	94.36	94.4	92.85	92.85	93.35	94.15	93.03	93.03	93.03	93.03	93.67	93.62	93.52	93.72
CHN	92.39		94.2	94.26	94.47	94.89	95.08	94.34	94.34	94.89	94.89	94.04	94.04	94.04	95.64	95.73	93.8	93.8	94.57	95.32	94.26	94.26	94.26	94.26	94.73	94.15	94.11	94.57
VN	93.35	94.2		97.66	97.98	95.32	94.55	94.39	94.39	95.16	95.16	94.79	94.79	94.79	96.28	96.32	94.55	94.55	94.95	95.37	94.79	94.79	94.79	94.79	94.84	95.53	95.12	95.16
VN	93.14	94.26	97.66		97.5	95.16	94.44	94.12	94.12	95.21	95.21	94.2	94.2	94.2	96.17	96.1	94.76	94.76	95.05	95.16	94.68	94.68	94.68	94.68	94.68	95.27	94.91	94.84
VN	93.64	94.47	97.98	97.5		95.53	94.92	94.34	94.34	95.43	95.43	94.73	94.73	94.73	96.65	96.58	94.71	94.71	95.16	95.32	95.32	95.32	95.32	95.32	94.89	95.69	95.28	95.21
VN	93.35	94.89	95.32	95.16	95.53		95.03	95.13	95.13	95.8	95.8	94.52	94.52	94.52	96.49	96.8	95.24	95.24	95.27	95.96	95.05	95.05	95.05	95.05	95.11	95.85	94.8	95.37
VN	93.32	95.08	94.55	94.44	94.92	95.03		95.27	95.27	95.66	95.66	94.28	94.28	94.28	95.98	96.61	94.68	94.68	95.03	95.82	94.71	94.71	94.71	94.71	94.55	94.97	94.83	95.08
U.S.	93.32	94.34	94.39	94.12	94.34	95.13	95.27		99.87	95.45	95.45	94.39	94.39	94.39	95.82	97.06	94.41	94.41	95.19	95.45	94.44	94.44	94.44	94.44	95.03	94.97	94.67	94.97
U.S.	93.32	94.34	94.39	94.12	94.34	95.13	95.27	99.87		95.45	95.45	94.39	94.39	94.39	95.82	97.06	94.41	94.41	95.19	95.45	94.44	94.44	94.44	94.44	95.03	94.97	94.67	94.97
U.S.	93.72	94.89	95.16	95.21	95.43	95.8	95.66	95.45	95.45		100	94.79	94.79	94.79	96.65	97.22	95.03	95.03	96.12	96.01	95.05	95.05	95.05	95.05	95.37	96.12	94.91	95.59
U.S.	93.72	94.89	95.16	95.21	95.43	95.8	95.66	95.45	95.45	100		94.79	94.79	94.79	96.65	97.22	95.03	95.03	96.12	96.01	95.05	95.05	95.05	95.05	95.37	96.12	94.91	95.59
U.S.	92.87	94.04	94.79	94.2	94.73	94.52	94.28	94.39	94.39	94.79	94.79		100	100	95.69	96.64	93.7	93.7	95.11	95.32	94.2	94.2	94.2	94.2	94.41	94.95	94.43	95.21
U.S.	92.87	94.04	94.79	94.2	94.73	94.52	94.28	94.39	94.39	94.79	94.79	100		100	95.69	96.64	93.7	93.7	95.11	95.32	94.2	94.2	94.2	94.2	94.41	94.95	94.43	95.21
U.S.	92.87	94.04	94.79	94.2	94.73	94.52	94.28	94.39	94.39	94.79	94.79	100	100		95.69	96.64	93.7	93.7	95.11	95.32	94.2	94.2	94.2	94.2	94.41	94.95	94.43	95.21
U.S.	94.36	95.64	96.28	96.17	96.65	96.49	95.98	95.82	95.82	96.65	96.65	95.69	95.69	95.69		98.02	96.04	96.04	96.6	96.6	95.9	95.9	95.9	95.9	96.38	96.81	95.76	96.33
U.S.	94.4	95.73	96.32	96.1	96.58	96.8	96.61	97.06	97.06	97.22	97.22	96.64	96.64	96.64	98.02		96.34	96.34	97.06	97.01	97.06	97.06	97.06	97.06	96.53	97.06	96.17	96.69
U.S.	92.85	93.8	94.55	94.76	94.71	95.24	94.68	94.41	94.41	95.03	95.03	93.7	93.7	93.7	96.04	96.34		99.81	95.08	94.97	94.65	94.65	94.65	94.65	94.71	94.34	94.61	94.44
U.S.	92.85	93.8	94.55	94.76	94.71	95.24	94.68	94.41	94.41	95.03	95.03	93.7	93.7	93.7	96.04	96.34	99.81		95.08	94.97	94.65	94.65	94.65	94.65	94.71	94.34	94.61	94.44
U.S.	93.35	94.57	94.95	95.05	95.16	95.27	95.03	95.19	95.19	96.12	96.12	95.11	95.11	95.11	96.6	97.06	95.08	95.08		96.01	95.48	95.48	95.48	95.48	95.48	96.06	94.75	95.59
U.S.	94.15	95.32	95.37	95.16	95.32	95.96	95.82	95.45	95.45	96.01	96.01	95.32	95.32	95.32	96.6	97.01	94.97	94.97	96.01		95.43	95.43	95.43	95.43	95.48	95.69	94.91	96.06
U.S.	93.03	94.26	94.79	94.68	95.32	95.05	94.71	94.44	94.44	95.05	95.05	94.2	94.2	94.2	95.9	97.06	94.65	94.65	95.48	95.43		100	100	100	95.21	95.48	94.59	94.63
U.S.	93.03	94.26	94.79	94.68	95.32	95.05	94.71	94.44	94.44	95.05	95.05	94.2	94.2	94.2	95.9	97.06	94.65	94.65	95.48	95.43	100		100	100	95.21	95.48	94.59	94.63
U.S.	93.03	94.26	94.79	94.68	95.32	95.05	94.71	94.44	94.44	95.05	95.05	94.2	94.2	94.2	95.9	97.06	94.65	94.65	95.48	95.43	100	100		100	95.21	95.48	94.59	94.63
U.S.	93.03	94.26	94.79	94.68	95.32	95.05	94.71	94.44	94.44	95.05	95.05	94.2	94.2	94.2	95.9	97.06	94.65	94.65	95.48	95.43	100	100	100		95.21	95.48	94.59	94.63
U.S.	93.67	94.73	94.84	94.68	94.89	95.11	94.55	95.03	95.03	95.37	95.37	94.41	94.41	94.41	96.38	96.53	94.71	94.71	95.48	95.48	95.21	95.21	95.21	95.21		95.16	94.32	94.89
UK	93.62	94.15	95.53	95.27	95.69	95.85	94.97	94.97	94.97	96.12	96.12	94.95	94.95	94.95	96.81	97.06	94.34	94.34	96.06	95.69	95.48	95.48	95.48	95.48	95.16		95.23	95.43
UK	93.52	94.11	95.12	94.91	95.28	94.8	94.83	94.67	94.67	94.91	94.91	94.43	94.43	94.43	95.76	96.17	94.61	94.61	94.75	94.91	94.59	94.59	94.59	94.59	94.32	95.23		94.59
U.S.	93.72	94.57	95.16	94.84	95.21	95.37	95.08	94.97	94.97	95.59	95.59	95.21	95.21	95.21	96.33	96.69	94.44	94.44	95.59	96.06	94.63	94.63	94.63	94.63	94.89	95.43	94.59	

NS4A	CHN	CHN	VN	VN	VN	VN	VN	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	UK	UK	U.S
CHN		94.31	95.12	95.93	94.31	95.93	93.5	93.5	93.5	95.12	95.12	94.31	94.31	94.31	95.93	96.75	92.68	92.68	95.93	95.93	95.12	95.12	95.12	95.12	95.93	94.31	88.62	94.31
CHN	94.31		97.56	96.75	93.5	93.5	92.68	90.24	90.24	94.31	94.31	95.12	95.12	95.12	94.31	93.5	92.68	92.68	95.12	94.31	92.68	92.68	92.68	92.68	95.12	92.68	86.99	94.31
VN	95.12	97.56		99.19	95.93	92.68	93.5	91.87	91.87	95.93	95.93	95.93	95.93	95.93	95.93	95.12	92.68	92.68	97.56	95.93	93.5	93.5	93.5	93.5	97.56	94.31	88.62	95.93
VN	95.93	96.75	99.19		96.75	93.5	94.31	92.68	92.68	96.75	96.75	95.12	95.12	95.12	96.75	95.93	93.5	93.5	96.75	96.75	94.31	94.31	94.31	94.31	98.37	95.12	87.8	96.75
VN	94.31	93.5	95.93	96.75		91.87	92.68	91.87	91.87	95.12	95.12	91.87	91.87	91.87	94.31	94.31	90.24	90.24	95.12	95.12	92.68	92.68	92.68	92.68	95.12	93.5	86.18	93.5
VN	95.93	93.5	92.68	93.5	91.87		95.93	91.87	91.87	93.5	93.5	93.5	93.5	93.5	92.68	93.5	89.43	89.43	91.87	94.31	92.68	92.68	92.68	92.68	95.12	94.31	85.37	94.31
VN	93.5	92.68	93.5	94.31	92.68	95.93		94.31	94.31	94.31	94.31	94.31	94.31	94.31	93.5	94.31	90.24	90.24	92.68	95.12	91.87	91.87	91.87	91.87	95.93	95.12	86.18	95.12
U.S.	93.5	90.24	91.87	92.68	91.87	91.87	94.31		100	94.31	94.31	95.12	95.12	95.12	95.12	95.93	91.06	91.06	92.68	93.5	91.87	91.87	91.87	91.87	94.31	91.87	86.99	91.87
U.S.	93.5	90.24	91.87	92.68	91.87	91.87	94.31	100		94.31	94.31	95.12	95.12	95.12	95.12	95.93	91.06	91.06	92.68	93.5	91.87	91.87	91.87	91.87	94.31	91.87	86.99	91.87
U.S.	95.12	94.31	95.93	96.75	95.12	93.5	94.31	94.31	94.31		100	95.12	95.12	95.12	95.93	98.37	92.68	92.68	95.12	97.56	95.12	95.12	95.12	95.12	95.93	95.93	88.62	95.93
U.S.	95.12	94.31	95.93	96.75	95.12	93.5	94.31	94.31	94.31	100		95.12	95.12	95.12	95.93	98.37	92.68	92.68	95.12	97.56	95.12	95.12	95.12	95.12	95.93	95.93	88.62	95.93
U.S.	94.31	95.12	95.93	95.12	91.87	93.5	94.31	95.12	95.12	95.12	95.12		100	100	95.93	96.75	92.68	92.68	95.12	94.31	92.68	92.68	92.68	92.68	96.75	92.68	89.43	94.31
U.S.	94.31	95.12	95.93	95.12	91.87	93.5	94.31	95.12	95.12	95.12	95.12	100		100	95.93	96.75	92.68	92.68	95.12	94.31	92.68	92.68	92.68	92.68	96.75	92.68	89.43	94.31
U.S.	94.31	95.12	95.93	95.12	91.87	93.5	94.31	95.12	95.12	95.12	95.12	100	100		95.93	96.75	92.68	92.68	95.12	94.31	92.68	92.68	92.68	92.68	96.75	92.68	89.43	94.31
U.S.	95.93	94.31	95.93	96.75	94.31	92.68	93.5	95.12	95.12	95.93	95.93	95.93	95.93	95.93		97.56	95.12	95.12	95.12	96.75	92.68	92.68	92.68	92.68	96.75	95.12	88.62	96.75
U.S.	96.75	93.5	95.12	95.93	94.31	93.5	94.31	95.93	95.93	98.37	98.37	96.75	96.75	96.75	97.56		94.31	94.31	95.93	97.56	95.12	95.12	95.12	95.12	95.93	95.93	90.24	95.93
U.S.	92.68	92.68	92.68	93.5	90.24	89.43	90.24	91.06	91.06	92.68	92.68	92.68	92.68	92.68	95.12	94.31		100	92.68	93.5	89.43	89.43	89.43	89.43	93.5	91.87	88.62	93.5
U.S.	92.68	92.68	92.68	93.5	90.24	89.43	90.24	91.06	91.06	92.68	92.68	92.68	92.68	92.68	95.12	94.31	100		92.68	93.5	89.43	89.43	89.43	89.43	93.5	91.87	88.62	93.5
U.S.	95.93	95.12	97.56	96.75	95.12	91.87	92.68	92.68	92.68	95.12	95.12	95.12	95.12	95.12	95.93	92.68	92.68		95.12	92.68	92.68	92.68	92.68	95.12	93.5	90.24	93.5
U.S.	95.93	94.31	95.93	96.75	95.12	94.31	95.12	93.5	93.5	97.56	97.56	94.31	94.31	94.31	96.75	97.56	93.5	93.5	95.12		94.31	94.31	94.31	94.31	96.75	96.75	88.62	96.75
U.S.	95.12	92.68	93.5	94.31	92.68	92.68	91.87	91.87	91.87	95.12	95.12	92.68	92.68	92.68	92.68	95.12	89.43	89.43	92.68	94.31		100	100	100	94.31	92.68	86.18	92.68
U.S.	95.12	92.68	93.5	94.31	92.68	92.68	91.87	91.87	91.87	95.12	95.12	92.68	92.68	92.68	92.68	95.12	89.43	89.43	92.68	94.31	100		100	100	94.31	92.68	86.18	92.68
U.S.	95.12	92.68	93.5	94.31	92.68	92.68	91.87	91.87	91.87	95.12	95.12	92.68	92.68	92.68	92.68	95.12	89.43	89.43	92.68	94.31	100	100		100	94.31	92.68	86.18	92.68
U.S.	95.12	92.68	93.5	94.31	92.68	92.68	91.87	91.87	91.87	95.12	95.12	92.68	92.68	92.68	92.68	95.12	89.43	89.43	92.68	94.31	100	100	100		94.31	92.68	86.18	92.68
U.S.	95.93	95.12	97.56	98.37	95.12	95.12	95.93	94.31	94.31	95.93	95.93	96.75	96.75	96.75	96.75	95.93	93.5	93.5	95.12	96.75	94.31	94.31	94.31	94.31		95.12	87.8	96.75
UK	94.31	92.68	94.31	95.12	93.5	94.31	95.12	91.87	91.87	95.93	95.93	92.68	92.68	92.68	95.12	95.93	91.87	91.87	93.5	96.75	92.68	92.68	92.68	92.68	95.12		88.62	96.75
UK	88.62	86.99	88.62	87.8	86.18	85.37	86.18	86.99	86.99	88.62	88.62	89.43	89.43	89.43	88.62	90.24	88.62	88.62	90.24	88.62	86.18	86.18	86.18	86.18	87.8	88.62		88.62
U.S.	94.31	94.31	95.93	96.75	93.5	94.31	95.12	91.87	91.87	95.93	95.93	94.31	94.31	94.31	96.75	95.93	93.5	93.5	93.5	96.75	92.68	92.68	92.68	92.68	96.75	96.75	88.62	

NS4B	CHN	CHN	VN	VN	VN	VN	VN	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	UK	UK	U.S
CHN		92.46	91.82	92.39	92.01	92.71	91.62	92.97	92.97	92.46	92.46	92.84	92.84	92.84	93.86	93.86	92.01	92.01	93.22	93.22	92.33	92.33	92.33	92.33	93.73	93.22	92.07	93.73
CHN	92.46		94.5	93.8	94.18	94.5	93.29	93.35	93.35	93.73	93.73	94.76	94.76	94.76	95.78	96.04	93.54	93.54	94.88	95.4	93.48	93.48	93.48	93.48	93.99	95.4	93.86	94.25
VN	91.82	94.5		97.12	96.61	94.37	92.77	93.61	93.61	94.12	94.12	93.73	93.73	93.73	95.27	94.76	92.65	92.65	94.12	93.86	92.07	92.07	92.07	92.07	94.25	93.86	92.46	93.61
VN	92.39	93.8	97.12		96.8	94.69	92.71	93.54	93.54	93.67	93.67	93.93	93.93	93.93	94.69	94.57	92.58	92.58	94.31	94.57	92.01	92.01	92.01	92.01	94.57	94.57	92.52	93.41
VN	92.01	94.18	96.61	96.8		93.8	92.33	93.41	93.41	93.16	93.16	93.16	93.16	93.16	94.31	94.05	91.94	91.94	93.29	93.67	91.11	91.11	91.11	91.11	93.41	94.18	91.75	93.29
VN	92.71	94.5	94.37	94.69	93.8		93.03	93.09	93.09	94.5	94.5	94.88	94.88	94.88	95.14	95.52	93.29	93.29	95.27	94.88	92.58	92.58	92.58	92.58	94.25	94.76	93.22	94.25
VN	91.62	93.29	92.77	92.71	92.33	93.03		92.14	92.14	92.52	92.52	92.77	92.77	92.77	94.31	93.93	92.33	92.33	93.03	93.16	91.62	91.62	91.62	91.62	93.41	94.18	90.98	92.9
U.S.	92.97	93.35	93.61	93.54	93.41	93.09	92.14		100	92.97	92.97	92.97	92.97	92.97	94.63	94.88	92.77	92.77	93.09	94.12	91.56	91.56	91.56	91.56	93.22	93.35	92.84	93.35
U.S.	92.97	93.35	93.61	93.54	93.41	93.09	92.14	100		92.97	92.97	92.97	92.97	92.97	94.63	94.88	92.77	92.77	93.09	94.12	91.56	91.56	91.56	91.56	93.22	93.35	92.84	93.35
U.S.	92.46	93.73	94.12	93.67	93.16	94.5	92.52	92.97	92.97		100	94.37	94.37	94.37	95.4	95.27	93.29	93.29	93.73	94.25	92.33	92.33	92.33	92.33	93.99	94.25	93.09	94.25
U.S.	92.46	93.73	94.12	93.67	93.16	94.5	92.52	92.97	92.97	100		94.37	94.37	94.37	95.4	95.27	93.29	93.29	93.73	94.25	92.33	92.33	92.33	92.33	93.99	94.25	93.09	94.25
U.S.	92.84	94.76	93.73	93.93	93.16	94.88	92.77	92.97	92.97	94.37	94.37		100	100	96.42	97.06	94.18	94.18	94.5	94.76	93.09	93.09	93.09	93.09	94.5	95.01	93.86	94.5
U.S.	92.84	94.76	93.73	93.93	93.16	94.88	92.77	92.97	92.97	94.37	94.37	100		100	96.42	97.06	94.18	94.18	94.5	94.76	93.09	93.09	93.09	93.09	94.5	95.01	93.86	94.5
U.S.	92.84	94.76	93.73	93.93	93.16	94.88	92.77	92.97	92.97	94.37	94.37	100	100		96.42	97.06	94.18	94.18	94.5	94.76	93.09	93.09	93.09	93.09	94.5	95.01	93.86	94.5
U.S.	93.86	95.78	95.27	94.69	94.31	95.14	94.31	94.63	94.63	95.4	95.4	96.42	96.42	96.42		97.44	94.57	94.57	95.78	95.78	94.37	94.37	94.37	94.37	95.78	95.78	94.5	96.04
U.S.	93.86	96.04	94.76	94.57	94.05	95.52	93.93	94.88	94.88	95.27	95.27	97.06	97.06	97.06	97.44		96.07	96.07	95.65	96.16	95.14	95.14	95.14	95.14	95.27	96.16	94.88	96.04
U.S.	92.01	93.54	92.65	92.58	91.94	93.29	92.33	92.77	92.77	93.29	93.29	94.18	94.18	94.18	94.57	96.07		99.42	93.16	94.18	92.65	92.65	92.65	92.65	93.16	93.67	93.03	93.67
U.S.	92.01	93.54	92.65	92.58	91.94	93.29	92.33	92.77	92.77	93.29	93.29	94.18	94.18	94.18	94.57	96.07	99.42		93.16	94.18	92.65	92.65	92.65	92.65	93.16	93.67	93.03	93.67
U.S.	93.22	94.88	94.12	94.31	93.29	95.27	93.03	93.09	93.09	93.73	93.73	94.5	94.5	94.5	95.78	95.65	93.16	93.16		95.52	93.73	93.73	93.73	93.73	95.65	95.65	93.86	94.88
U.S.	93.22	95.4	93.86	94.57	93.67	94.88	93.16	94.12	94.12	94.25	94.25	94.76	94.76	94.76	95.78	96.16	94.18	94.18	95.52		93.86	93.86	93.86	93.86	95.01	95.52	94.5	95.01
U.S.	92.33	93.48	92.07	92.01	91.11	92.58	91.62	91.56	91.56	92.33	92.33	93.09	93.09	93.09	94.37	95.14	92.65	92.65	93.73	93.86		100	100	100	93.48	93.86	93.73	92.97
U.S.	92.33	93.48	92.07	92.01	91.11	92.58	91.62	91.56	91.56	92.33	92.33	93.09	93.09	93.09	94.37	95.14	92.65	92.65	93.73	93.86	100		100	100	93.48	93.86	93.73	92.97
U.S.	92.33	93.48	92.07	92.01	91.11	92.58	91.62	91.56	91.56	92.33	92.33	93.09	93.09	93.09	94.37	95.14	92.65	92.65	93.73	93.86	100	100		100	93.48	93.86	93.73	92.97
U.S.	92.33	93.48	92.07	92.01	91.11	92.58	91.62	91.56	91.56	92.33	92.33	93.09	93.09	93.09	94.37	95.14	92.65	92.65	93.73	93.86	100	100	100		93.48	93.86	93.73	92.97
U.S.	93.73	93.99	94.25	94.57	93.41	94.25	93.41	93.22	93.22	93.99	93.99	94.5	94.5	94.5	95.78	95.27	93.16	93.16	95.65	95.01	93.48	93.48	93.48	93.48		95.01	93.61	95.01
UK	93.22	95.4	93.86	94.57	94.18	94.76	94.18	93.35	93.35	94.25	94.25	95.01	95.01	95.01	95.78	96.16	93.67	93.67	95.65	95.52	93.86	93.86	93.86	93.86	95.01		94.12	95.27
UK	92.07	93.86	92.46	92.52	91.75	93.22	90.98	92.84	92.84	93.09	93.09	93.86	93.86	93.86	94.5	94.88	93.03	93.03	93.86	94.5	93.73	93.73	93.73	93.73	93.61	94.12		94.37
U.S.	93.73	94.25	93.61	93.41	93.29	94.25	92.9	93.35	93.35	94.25	94.25	94.5	94.5	94.5	96.04	96.04	93.67	93.67	94.88	95.01	92.97	92.97	92.97	92.97	95.01	95.27	94.37	

NS5A	CHN	CHN	VN	VN	VN	VN	VN	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	UK	UK	U.S
CHN		92.63	90.58	91.9	91.72	91.35	92.19	91.39	91.39	92.03	92.03	92.04	92.04	92.04	94.53	93.96	92.34	92.34	92.19	91.61	90.95	90.95	90.95	90.95	92.08	90.66	89.58	90.8
CHN	92.63		91.68	92.48	92.74	92.88	93.28	93.07	93.07	93.41	93.41	92.85	92.85	92.85	95.47	95.05	93.07	93.07	93.36	92.99	92.19	92.19	92.19	92.19	93.54	91.68	90.16	92.48
VN	90.58	91.68		94.96	95	91.5	91.31	91.02	91.02	91.66	91.66	91.09	91.09	91.09	93.07	92.79	91.46	91.46	91.2	91.02	90	90	90	90	92.23	89.85	88.92	90.58
VN	91.9	92.48	94.96		96.39	91.93	92.7	91.75	91.75	92.54	92.54	92.34	92.34	92.34	94.67	93.74	92.55	92.55	92.52	92.55	90.95	90.95	90.95	90.95	93.1	91.46	89.8	91.53
VN	91.72	92.74	95	96.39		92.77	93.21	92.23	92.23	92.72	92.72	92.74	92.74	92.74	94.64	94.22	93.18	93.18	93.36	92.3	91.28	91.28	91.28	91.28	93.36	91.57	90.49	91.93
VN	91.35	92.88	91.5	91.93	92.77		92.37	92.08	92.08	92.14	92.14	92.3	92.3	92.3	94.05	94.29	92.74	92.74	92.52	91.93	91.64	91.64	91.64	91.64	92.7	90.91	89.98	90.91
VN	92.19	93.28	91.31	92.7	93.21	92.37		92.96	92.96	92.68	92.68	92.15	92.15	92.15	94.82	94.76	93.36	93.36	93.69	91.82	92.08	92.08	92.08	92.08	93.32	91.93	90.6	91.9
U.S.	91.39	93.07	91.02	91.75	92.23	92.08	92.96		100	92.83	92.83	92.19	92.19	92.19	94.53	95.27	93.28	93.28	93.43	91.68	91.46	91.46	91.46	91.46	93.32	91.24	90.24	91.9
U.S.	91.39	93.07	91.02	91.75	92.23	92.08	92.96	100		92.83	92.83	92.19	92.19	92.19	94.53	95.27	93.28	93.28	93.43	91.68	91.46	91.46	91.46	91.46	93.32	91.24	90.24	91.9
U.S.	92.03	93.41	91.66	92.54	92.72	92.14	92.68	92.83	92.83		99.84	92.61	92.61	92.61	95.02	94.67	93.49	93.49	92.68	92.46	91.37	91.37	91.37	91.37	93.3	91.44	89.56	91.51
U.S.	92.03	93.41	91.66	92.54	92.72	92.14	92.68	92.83	92.83	99.84		92.61	92.61	92.61	95.02	94.67	93.49	93.49	92.68	92.46	91.37	91.37	91.37	91.37	93.3	91.44	89.56	91.51
U.S.	92.04	92.85	91.09	92.34	92.74	92.3	92.15	92.19	92.19	92.61	92.61		100	100	94.82	96.15	93.8	93.8	92.85	92.92	92.12	92.12	92.12	92.12	92.59	90.8	90.31	92.12
U.S.	92.04	92.85	91.09	92.34	92.74	92.3	92.15	92.19	92.19	92.61	92.61	100		100	94.82	96.15	93.8	93.8	92.85	92.92	92.12	92.12	92.12	92.12	92.59	90.8	90.31	92.12
U.S.	92.04	92.85	91.09	92.34	92.74	92.3	92.15	92.19	92.19	92.61	92.61	100	100		94.82	96.15	93.8	93.8	92.85	92.92	92.12	92.12	92.12	92.12	92.59	90.8	90.31	92.12
U.S.	94.53	95.47	93.07	94.67	94.64	94.05	94.82	94.53	94.53	95.02	95.02	94.82	94.82	94.82		97.32	95.69	95.69	94.96	94.16	93.28	93.28	93.28	93.28	95.22	92.92	91.48	94.16
U.S.	93.96	95.05	92.79	93.74	94.22	94.29	94.76	95.27	95.27	94.67	94.67	96.15	96.15	96.15	97.32		96.81	96.81	94.69	94.11	94.84	94.84	94.84	94.84	95.24	92.79	91.72	93.67
U.S.	92.34	93.07	91.46	92.55	93.18	92.74	93.36	93.28	93.28	93.49	93.49	93.8	93.8	93.8	95.69	96.81		100	93.36	92.7	92.34	92.34	92.34	92.34	93.47	91.53	90.24	91.97
U.S.	92.34	93.07	91.46	92.55	93.18	92.74	93.36	93.28	93.28	93.49	93.49	93.8	93.8	93.8	95.69	96.81	100		93.36	92.7	92.34	92.34	92.34	92.34	93.47	91.53	90.24	91.97
U.S.	92.19	93.36	91.2	92.52	93.36	92.52	93.69	93.43	93.43	92.68	92.68	92.85	92.85	92.85	94.96	94.69	93.36	93.36		92.55	91.46	91.46	91.46	91.46	93.03	91.61	91.11	91.82
U.S.	91.61	92.99	91.02	92.55	92.3	91.93	91.82	91.68	91.68	92.46	92.46	92.92	92.92	92.92	94.16	94.11	92.7	92.7	92.55		91.24	91.24	91.24	91.24	92.52	90.88	89.51	90.88
U.S.	90.95	92.19	90	90.95	91.28	91.64	92.08	91.46	91.46	91.37	91.37	92.12	92.12	92.12	93.28	94.84	92.34	92.34	91.46	91.24		100	100	100	92.37	89.93	89.58	90.44
U.S.	90.95	92.19	90	90.95	91.28	91.64	92.08	91.46	91.46	91.37	91.37	92.12	92.12	92.12	93.28	94.84	92.34	92.34	91.46	91.24	100		100	100	92.37	89.93	89.58	90.44
U.S.	90.95	92.19	90	90.95	91.28	91.64	92.08	91.46	91.46	91.37	91.37	92.12	92.12	92.12	93.28	94.84	92.34	92.34	91.46	91.24	100	100		100	92.37	89.93	89.58	90.44
U.S.	90.95	92.19	90	90.95	91.28	91.64	92.08	91.46	91.46	91.37	91.37	92.12	92.12	92.12	93.28	94.84	92.34	92.34	91.46	91.24	100	100	100		92.37	89.93	89.58	90.44
U.S.	92.08	93.54	92.23	93.1	93.36	92.7	93.32	93.32	93.32	93.3	93.3	92.59	92.59	92.59	95.22	95.24	93.47	93.47	93.03	92.52	92.37	92.37	92.37	92.37		92.3	90.93	92.3
UK	90.66	91.68	89.85	91.46	91.57	90.91	91.93	91.24	91.24	91.44	91.44	90.8	90.8	90.8	92.92	92.79	91.53	91.53	91.61	90.88	89.93	89.93	89.93	89.93	92.3		89.14	90.15
UK	89.58	90.16	88.92	89.8	90.49	89.98	90.6	90.24	90.24	89.56	89.56	90.31	90.31	90.31	91.48	91.72	90.24	90.24	91.11	89.51	89.58	89.58	89.58	89.58	90.93	89.14		89.14
U.S.	90.8	92.48	90.58	91.53	91.93	90.91	91.9	91.9	91.9	91.51	91.51	92.12	92.12	92.12	94.16	93.67	91.97	91.97	91.82	90.88	90.44	90.44	90.44	90.44	92.3	90.15	89.14	

NS5B	CHN	CHN	VN	VN	VN	VN	VN	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	U.S	UK	UK	U.S
CHN		92.51	91.86	92.21	92.04	92.51	92.68	92.51	92.51	93.03	93.03	92.45	92.51	92.45	93.44	94.01	92.51	92.51	92.77	92.21	93.09	93.09	93.09	93.09	92.45	92.45	93.11	93.03
CHN	92.51		94.09	94.26	93.79	93.68	94.32	93.5	93.5	93.85	93.85	93.44	93.5	93.44	94.91	95.36	93.68	93.68	93.94	93.44	94.38	94.38	94.38	94.38	93.44	93.15	93.93	93.68
VN	91.86	94.09		97.25	97.48	93.79	93.5	93.85	93.85	94.56	94.56	93.85	93.91	93.85	94.96	95.48	93.97	93.97	93.76	94.15	94.5	94.5	94.5	94.5	93.79	92.92	94.04	93.79
VN	92.21	94.26	97.25		97.37	94.26	93.62	94.15	94.15	94.5	94.5	94.2	94.26	94.2	95.14	96	94.35	94.35	94.06	94.44	94.79	94.79	94.79	94.79	94.26	93.27	94.22	94.32
VN	92.04	93.79	97.48	97.37		93.74	93.68	93.74	93.74	94.44	94.44	93.97	94.03	93.97	95.14	95.71	94.12	94.12	94.17	94.38	94.44	94.44	94.44	94.44	93.79	93.03	94.16	93.97
VN	92.51	93.68	93.79	94.26	93.74		93.15	93.15	93.15	93.97	93.97	93.27	93.33	93.27	94.91	95.3	94.17	94.17	94	93.74	94.5	94.5	94.5	94.5	93.38	94.09	93.93	94.38
VN	92.68	94.32	93.5	93.62	93.68	93.15		93.56	93.56	93.62	93.62	93.03	93.09	93.03	95.02	95.18	94.26	94.26	94.12	93.15	94.38	94.38	94.38	94.38	94.03	93.15	94.04	93.33
U.S.	92.51	93.5	93.85	94.15	93.74	93.15	93.56		100	94.44	94.44	94.38	94.44	94.38	95.26	96.53	94.76	94.76	94.29	93.5	94.26	94.26	94.26	94.26	94.09	93.21	94.34	93.97
U.S.	92.51	93.5	93.85	94.15	93.74	93.15	93.56	100		94.44	94.44	94.38	94.44	94.38	95.26	96.53	94.76	94.76	94.29	93.5	94.26	94.26	94.26	94.26	94.09	93.21	94.34	93.97
U.S.	93.03	93.85	94.56	94.5	94.44	93.97	93.62	94.44	94.44		100	94.5	94.56	94.5	95.49	96.3	94.79	94.79	94.41	94.38	94.85	94.85	94.85	94.85	94.03	93.74	94.57	94.38
U.S.	93.03	93.85	94.56	94.5	94.44	93.97	93.62	94.44	94.44	100		94.5	94.56	94.5	95.49	96.3	94.79	94.79	94.41	94.38	94.85	94.85	94.85	94.85	94.03	93.74	94.57	94.38
U.S.	92.45	93.44	93.85	94.2	93.97	93.27	93.03	94.38	94.38	94.5	94.5		99.94	100	95.02	96.59	94.12	94.12	94.12	93.85	94.44	94.44	94.44	94.44	94.61	92.45	94.28	93.79
U.S.	92.51	93.5	93.91	94.26	94.03	93.33	93.09	94.44	94.44	94.56	94.56	99.94		99.94	95.08	96.65	94.17	94.17	94.17	93.91	94.5	94.5	94.5	94.5	94.67	92.51	94.34	93.85
U.S.	92.45	93.44	93.85	94.2	93.97	93.27	93.03	94.38	94.38	94.5	94.5	100	99.94		95.02	96.59	94.12	94.12	94.12	93.85	94.44	94.44	94.44	94.44	94.61	92.45	94.28	93.79
U.S.	93.44	94.91	94.96	95.14	95.14	94.91	95.02	95.26	95.26	95.49	95.49	95.02	95.08	95.02		97.41	95.78	95.78	95.99	95.32	95.96	95.96	95.96	95.96	95.73	94.38	95.1	95.49
U.S.	94.01	95.36	95.48	96	95.71	95.3	95.18	96.53	96.53	96.3	96.3	96.59	96.65	96.59	97.41		96.56	96.56	96.27	95.65	96.88	96.88	96.88	96.88	95.95	94.77	95.9	95.77
U.S.	92.51	93.68	93.97	94.35	94.12	94.17	94.26	94.76	94.76	94.79	94.79	94.12	94.17	94.12	95.78	96.56		99.85	94.58	94.29	94.47	94.47	94.47	94.47	94.41	93.62	94.6	94.06
U.S.	92.51	93.68	93.97	94.35	94.12	94.17	94.26	94.76	94.76	94.79	94.79	94.12	94.17	94.12	95.78	96.56	99.85		94.58	94.29	94.47	94.47	94.47	94.47	94.41	93.62	94.6	94.06
U.S.	92.77	93.94	93.76	94.06	94.17	94	94.12	94.29	94.29	94.41	94.41	94.12	94.17	94.12	95.99	96.27	94.58	94.58		94.76	94.88	94.88	94.88	94.88	94.53	93.76	94.83	94.47
U.S.	92.21	93.44	94.15	94.44	94.38	93.74	93.15	93.5	93.5	94.38	94.38	93.85	93.91	93.85	95.32	95.65	94.29	94.29	94.76		94.85	94.85	94.85	94.85	94.44	93.79	94.22	93.91
U.S.	93.09	94.38	94.5	94.79	94.44	94.5	94.38	94.26	94.26	94.85	94.85	94.44	94.5	94.44	95.96	96.88	94.47	94.47	94.88	94.85		100	100	100	94.79	94.15	94.39	95.43
U.S.	93.09	94.38	94.5	94.79	94.44	94.5	94.38	94.26	94.26	94.85	94.85	94.44	94.5	94.44	95.96	96.88	94.47	94.47	94.88	94.85	100		100	100	94.79	94.15	94.39	95.43
U.S.	93.09	94.38	94.5	94.79	94.44	94.5	94.38	94.26	94.26	94.85	94.85	94.44	94.5	94.44	95.96	96.88	94.47	94.47	94.88	94.85	100	100		100	94.79	94.15	94.39	95.43
U.S.	93.09	94.38	94.5	94.79	94.44	94.5	94.38	94.26	94.26	94.85	94.85	94.44	94.5	94.44	95.96	96.88	94.47	94.47	94.88	94.85	100	100	100		94.79	94.15	94.39	95.43
U.S.	92.45	93.44	93.79	94.26	93.79	93.38	94.03	94.09	94.09	94.03	94.03	94.61	94.67	94.61	95.73	95.95	94.41	94.41	94.53	94.44	94.79	94.79	94.79	94.79		93.15	94.75	94.03
UK	92.45	93.15	92.92	93.27	93.03	94.09	93.15	93.21	93.21	93.74	93.74	92.45	92.51	92.45	94.38	94.77	93.62	93.62	93.76	93.79	94.15	94.15	94.15	94.15	93.15		93.22	93.33
UK	93.11	93.93	94.04	94.22	94.16	93.93	94.04	94.34	94.34	94.57	94.57	94.28	94.34	94.28	95.1	95.9	94.6	94.6	94.83	94.22	94.39	94.39	94.39	94.39	94.75	93.22		93.63
U.S.	93.03	93.68	93.79	94.32	93.97	94.38	93.33	93.97	93.97	94.38	94.38	93.79	93.85	93.79	95.49	95.77	94.06	94.06	94.47	93.91	95.43	95.43	95.43	95.43	94.03	93.33	93.63	

Appendix 9: dN/dS ratio. Note: NA: not available

Gene	S	E1	E2	X	NS2	NS3	NS4A	NS4B	NS5A	NS5B	Full CDS
Mean dN/dS (Vietnam)	NA	0.264	0.17	0.31	0.44	0.226	0.277	0.174	0.221	0.204	0.232
Mean dN/dS (US)	0.198	0.129	0.214	0.203	0.212	0.158	0.241	0.15	0.262	0.167	0.198

Appendix 10: Routine diagnostic workup during the study period at the study site

Suspected clinical entity	**First line diagnosis**	**Additional testing***
Tuberculous meningitis	Ziehl Neelsen stain, GenXpert and culture[#]	NA
Bacterial meningitis	Gram stain and culture	
Meningoencephalitis	HSV PCR	VZV PCR and serological testing for flaviviruses (including JEV and DENV), and mumps virus
Cryptococcus	Lateral Flow Assay	NA
Auto-immune encephalitis	ND	Anti-NMDAR encephalitis**

Note: *Up on requested by treating physicians, [#] using Mycobacteria Growth Indicator tubes, **retrospective testing

Appendix 11: Primers and probes used for PCR confirmatory testing

Viruses	Names of primer/probe	Oligo sequences (5'-3')	References
Enterovirus	ENT-F	CCCTGAATGCGGCTAAT	(232)
	ENT-R	ATTGTCACCATAAGCAGCC	
	ENTr-probe	Cy5-ACCCAAAGTAGTCGGTTCCG -BHQ3	
Mump virus	F1073	TCTCACCCATAGCAGGGAGTTATAT	(294)
	R1151	GTTAGACTTCGACAGTTTGCAACAA	
	Probe	FAM-AGGCGATTTGTA GCACTGGATG-TAMRA	
Rotavirus	NVP3-FDeg	ACCATCTWCACRTRACCCTC	(231)
	NVP3-R1	GGTCACATAACGCCCCTATA	
	NVP3-Probe	FAM-ATGAGCACAATAGTTAAAAGCTAACACTGTCAA-BHQ1	

Appendix 12: Results of routine diagnosis, expanded PCR testing and mNGS analysis

ID	ZN Smear	India Ink stain	Cryptococcal antigen test	Gram Stain	Bacterial culture	Others	HSV PCR (ct value)	JEV Serology	Dengue Serology	Dengue PCR	mNGS	Expanded PCR testing for EVs (ct value)	Expanded PCR testing for Mumps (Ct value)
S1	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	Negative	Negative	Negative
S2	Negative	Negative	ND	Negative	Negative	ND	Negative	ND	ND	ND	EVs	Negative	Negative
S3	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	Negative	Negative	Negative
S4	Negative	Negative	Negative	Negative	Negative	ND	ND	ND	ND	ND	Negative	Negative	Negative
S5	Negative	Negative	Negative	Negative	Negative	ND	Positive (25.1)	ND	ND	ND	HSV/EVs	Negative	Negative
S6	Negative	Negative	Negative	Negative	Negative	ND	Negative	Negative	Negative	ND	Negative	Negative	Negative
S7	Negative	Negative	Negative	Negative	Negative	ND	Negative	Negative	ND	ND	Negative	Negative	Negative
S8	Negative	ND	Negative	Negative	Negative	VZV PCR positive	Negative	ND	ND	ND	VZV/EVs	Negative	Negative
S9	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	EVs	Positive (33.36)	Negative
S10	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	EVs	Positive (34.25)	Negative
S11	Negative	Negative	Negative	Negative	Negative	ND	ND	ND	ND	ND	EVs	Negative	Negative
S12	Negative	Negative	ND	Negative	Negative	ND	Negative	ND	Negative	ND	EVs	Negative	Negative
S13	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	EVs	Negative	Negative
S14	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	Rotavirus	Negative	Negative
S15	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	EVs	Negative	Negative
S16	Negative	Negative	ND	Negative	Negative	ND	Negative	ND	ND	ND	Negative	Negative	Negative
S17	Negative	Negative	ND	Negative	Negative	ND	Negative	ND	ND	ND	Negative	Negative	Negative
S18	Negative	Negative	Negative	Negative	Negative	ND	ND	ND	ND	ND	Negative	Negative	Negative
S19	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	Negative	ND	EVs	Negative	Negative
S20	Negative	Negative	Negative	Negative	Negative	ND	ND	ND	ND	ND	Negative	Negative	Negative
S21	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	EVs	Negative	Negative
S22	Negative	ND	Negative	Negative	Negative	ND	ND	ND	ND	ND	EVs	Positive (34.79)	Negative
S23	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	Negative	ND	EVs	Negative	Negative
S24	Negative	Negative	ND	Negative	Negative	ND	Negative	Negative	ND	ND	Negative	Negative	Negative
S25	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	Negative	Negative	Negative
S26	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	Negative	Negative	Negative
S27	Negative	Negative	ND	Negative	ND	ND	Negative	Negative	Negative	ND	EVs	Negative	Negative
S28	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	Negative	Negative	Negative
S29	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	Negative	Negative	Negative
S30	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	EVs	Negative	Negative
S31	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	Negative	ND	Negative	Negative	Negative
S32	ND	Negative	ND	Negative	ND	ND	ND	ND	ND	ND	Negative	Negative	Negative
S33	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	Negative	ND	Negative	Negative	Negative
S34	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	Negative	ND	Negative	Negative	Negative
S35	Negative	Negative	Negative	Negative	Negative	ND	ND	ND	ND	ND	Negative	Negative	Negative

S36	Negative	Negative	Negative	Negative	Negative	ND	Positive (28.01)	ND	ND	ND	HSV	Negative	Negative
S37	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	Negative	Negative	Negative
S38	Negative	ND	Negative	Negative	Negative	ND	Negative	ND	ND	ND	EVs	Negative	Negative
S39	Negative	ND	Negative	Negative	Negative	ND	Negative	Negative	Positive	ND	Negative	Negative	Negative
S40	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	Negative	Negative	Negative
S41	Negative	Negative	Negative	Negative	Negative	Mumps virus IgG+IgM: Positive	Negative	ND	Negative	ND	Negative	Negative	Negative
S42	Negative	Negative	Negative	Negative	Negative	ND	ND	ND	ND	ND	EVs	Positive (34.78	Negative
S43	Negative	ND	Negative	Negative	Negative	ND	Negative	Negative	Negative	ND	EVs	Positive (31.23)	Negative
S44	ND	Negative	ND	Negative	ND	ND	ND	ND	ND	ND	Negative	Negative	Negative
S45	Negative	Negative	Negative	Negative	Negative	ND	Negative	Negative	ND	ND	Negative	Negative	Negative
S46	Negative	Negative	Negative	Negative	Negative	Mumps virus IgG+IgM: Positive	ND	ND	ND	ND	Mumps	Negative	positive (35.36)
S47	ND	Negative	ND	Negative	Negative	ND	ND	ND	Negative	ND	Negative	Negative	Negative
S48	ND	Negative	ND	Negative	Negative	ND	ND	ND	ND	ND	Negative	Negative	Negative
S49	Negative	ND	Negative	Negative	Negative	dengue rapid test: negative	Negative	Negative	ND	ND	EVs	positive (32.3)	Negative
S50	Negative	Negative	Negative	Negative	Negative	ND	ND	ND	ND	ND	EVs	Positive(32.65)	Negative
S51	Negative	Negative	Negative	Negative	Negative	ND	Negative	Negative	ND	ND	Negative	Negative	Negative
S52	Negative	Negative	Negative	Negative	Negative	Mumps virus IgG+IgM: Positive	ND	ND	ND	ND	Negative	Negative	positive (35.85)
S53	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	Negative	Negative	Negative
S54	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	Negative	ND	EVs	Negative	Negative
S55	ND	ND	ND	Negative	Negative	ND	Negative	ND	Positive	Negative	Negative	Negative	Negative
S56	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	Negative	Negative	Negative
S57	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	Negative	ND	Negative	Negative	Negative
S58	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	Negative	Negative	Negative
S59	Negative	Negative	Negative	Negative	Negative	ND	ND	ND	ND	ND	Negative	Negative	Negative
S60	Negative	Negative	Negative	Negative	Negative	ND	Negative	ND	ND	ND	Negative	Negative	Negative
S61	Negative	Negative	Negative	Negative	Negative	Mumps virus IgG+IgM: Positive	Negative	ND	ND	ND	Negative	Negative	positive (40)
S62	Negative	Negative	Negative	Negative	Negative	ND	Positive (unavailble)	ND	ND	ND	HSV/EVs	Negative	Negative

S63	ND	Negative	ND	Negative	ND	ND	positive (30.36)	Negative	ND	ND	HSV	negative	Negative
S64	Negative	Negative	Negative	Negative	Negative	ND	Positive (23.77)	ND	ND	ND	HSV	Negative	Negative
S65	Negative	Negative	Negative	Negative	Negative	ND	Positive (30.62)	ND	ND	ND	HSV	Negative	Negative
S66	Negative	ND	Negative	Negative	Negative	ND	Negative	ND	ND	ND	Negative	Negative	positive (40)
S67	Negative	Negative	Negative	Negative	Negative	ND	Negative	Positive	ND	ND	EVs	Negative	Negative
S68	Negative	Negative	Negative	Negative	Negative	ND	Positive 28.71)	ND	ND	ND	HSV	Negative	Negative

Note: * S19 & S26: Negative controls; ND: not done

Appendix 13: mNGS reads obtained from DNA/RNA-virus workflows

	Total reads	Median	Range
DNA-virus workflow	62,565,802	859,656	1,487,000 – 2,125,00
RNA-virus workflow	49,233,869	717,707	7,368 – 5,874,00

Appendix 14: EV reads of PCR negative samples identical to reads found in sample(s) with a high abundance of EV reads with which they (did not) share(d) an index

CSF number	**Number of unique EV reads found in PCR negative samples**	**Number of unique EV reads (%) (1)**	**Number of unique EV reads (%) (2)**
17	2	1 (50)	0
18	2	2 (100)	0
19	4	0	0
20	1	0	0
21	2	2 (100)	0
22	5	5 (100)	0
23	13	4 (31)	0
24	4	2 (50)	0
25	1	0	0
26	20	5 (25%)	0
27	7	4 (57)	0
28	24	12 (50)	0
29	22	0	0
30	20	2 (10)	0
31	12	7 (58)	0
32	4	2 (50)	0

Note:(1) identical to reads found in sample(s) with a high abundance of EV reads with which it shared an index. (2) identical to reads found in sample(s) with a high abundance of EV reads with which it did not shared an index.

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

VIROLOGY

Viruses in Vietnamese Patients Presenting with Community-Acquired Sepsis of Unknown Cause

Nguyen To Anh,[a] Nguyen Thi Thu Hong,[a] Le Nguyen Truc Nhu,[a] Tran Tan Thanh,[a] Chuen-Yen Lau,[c] Direk Limmathurotsakul,[b,d] Xutao Deng,[f,g] Motiur Rahman,[a] Nguyen Van Vinh Chau,[e] H. Rogier van Doorn,[a,b] Guy Thwaites,[a,b] Eric Delwart,[f,g] Le Van Tan,[a] for the Southeast Asia Infectious Disease Clinical Research Network

[a]Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam
[b]Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom
[c]Collaborative Clinical Research Branch, Division of Clinical Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA
[d]Mahidol Oxford Tropical Research Unit, Bangkok, Thailand
[e]Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam
[f]Blood Systems Research Institute, San Francisco, California, USA
[g]Department of Laboratory Medicine, University of California, San Francisco, California, USA

ABSTRACT Community-acquired (CA) sepsis is a major public health problem worldwide, yet the etiology remains unknown for >50% of the patients. Here we applied metagenomic next-generation sequencing (mNGS) to characterize the human virome in 492 clinical samples (384 sera, 92 pooled nasal and throat swabs, 10 stools, and 6 cerebrospinal fluid samples) from 386 patients (213 adults and 173 children) presenting with CA sepsis who were recruited from 6 hospitals across Vietnam between 2013 and 2015. Specific monoplex PCRs were used subsequently to confirm the presence of viral sequences detected by mNGS. We found sequences related to 47 viral species belonging to 21 families in 358 of 386 (93%) patients, including viruses known to cause human infections. After PCR confirmation, human viruses were found in 52 of 386 patients (13.4%); picornavirus (enteroviruses [$n = 14$], rhinovirus [$n = 5$], and parechovirus [$n = 2$]), hepatitis B virus ($n = 10$), cytomegalovirus ($n = 9$), Epstein-Barr virus ($n = 5$), and rotavirus A ($n = 3$) were the most common viruses detected. Recently discovered viruses were also found (gemycircularvirus [$n = 5$] and WU polyomavirus, Saffold virus, salivirus, cyclovirus-VN, and human pegivirus 2 [HPgV2] [n, 1 each]), adding to the growing literature about the geographic distribution of these novel viruses. Notably, sequences related to numerous viruses not previously reported in human tissues were also detected. To summarize, we identified 21 viral species known to be infectious to humans in 52 of 386 (13.4%) patients presenting with CA sepsis of unknown cause. The study, however, cannot directly impute sepsis causation to the viruses identified. The results highlight the fact that it remains a challenge to establish the causative agents in CA sepsis patients, especially in tropical settings such as Vietnam.

KEYWORDS Vietnam, community-acquired sepsis, viral metagenomics

According to the WHO, approximately 30 million cases of sepsis with 6 million deaths occur globally each year (1). Approximately 70% of sepsis cases are attributed to community-acquired (CA) infections (1). The increasing frequency of antimicrobial resistance and the diversity of pathogens (including bacteria and viruses) that may cause CA sepsis further complicate current diagnostic efforts, in turn posing challenges to patient management (2). Indeed, despite extensive laboratory investigations, the causes of a substantial proportion of cases of CA sepsis remain unknown. In

Citation Anh NT, Hong NTT, Nhu LNT, Thanh TT, Lau C-Y, Limmathurotsakul D, Deng X, Rahman M, Chau NVV, van Doorn HR, Thwaites G, Delwart E, Tan LV, for the Southeast Asia Infectious Disease Clinical Research Network. 2019. Viruses in Vietnamese patients presenting with community-acquired sepsis of unknown cause. J Clin Microbiol 57:e00386-19. https://doi.org/10.1128/JCM.00386-19.

Editor Yi-Wei Tang, Memorial Sloan Kettering Cancer Center

Address correspondence to Nguyen To Anh, anhnt@oucru.org, or Le Van Tan, tanlv@oucru.org.

Received 13 March 2019
Returned for modification 2 April 2019
Accepted 12 June 2019
Accepted manuscript posted online 19 June 2019
Published 26 August 2019

a recent etiological study of 1,578 patients with CA sepsis, conducted by the Southeast Asia Infectious Disease Clinical Research Network, the etiology (viruses, bacteria, and parasites) was established for only 48% (3). While this diagnostic yield is comparable to that of previous reports, the unknown etiology for >50% of the patients may be attributed to the low sensitivity of current diagnostic tests and/or the diversity of the causative agents that may be responsible for this important clinical condition. Furthermore, Southeast Asia is one of the major hot spots for the emergence of novel pathogens, as illustrated by the emergence of Nipah virus, severe acute respiratory syndrome (SARS) coronavirus, avian influenza virus A (H5N1), avian influenza virus A (H7N9), enterovirus A71 (EV-A71), and, more recently, Zika virus (4, 5).

Metagenomic next-generation sequencing (mNGS) has emerged as an unbiased, sequence-independent method for the detection of pathogens, especially viruses, in clinical samples (6–13). Using mNGS, we previously discovered a novel cyclovirus (cyclovirus-Vietnam [cyclovirus-VN]) in 4% of Vietnamese patients presenting with central nervous system (CNS) infections, although the pathogenicity and natural hosts of this virus remain unresolved (8, 14).

Improving our knowledge about the causative agents of CA sepsis can inform clinical management, while active surveillance for novel pathogens in this region is of public health significance. In this study, we use mNGS to characterize the viral contents of clinical samples collected from patients enrolled in an etiological study of sepsis of unknown etiology across Southeast Asia between 2013 and 2015 (3).

MATERIALS AND METHODS

Clinical specimens and patient data. The clinical specimens and patient data used for mNGS analysis were derived from an etiological study of CA sepsis conducted at multiple hospitals across Indonesia ($n = 3$), Thailand ($n = 4$), and Vietnam ($n = 6$) between 2013 and 2015 (3). Hospitalized patients with suspected or documented CA infections, fulfilling the diagnostic criteria for sepsis of the 2012 Surviving Sepsis Campaign (adults) (15) or the definitions of the Pediatric Sepsis Consensus Conference (16), were enrolled within 24 h of admission (3). A total of 1,582 patients were enrolled (750 each from Vietnam and Thailand; 82 from Indonesia) (Fig. 1). Per the study protocol, serum samples were collected from all patients; additional samples, including pooled nasal and throat swabs, cerebrospinal fluid (CSF), and stools, were collected when clinically indicated. After collection, all clinical samples were stored at −80°C. Additionally, information about the demographics, clinical entities, and outcomes of the patients was retrieved from a publicly available data set of the original study that was deposited at https://figshare.com/articles/Data_set_-_Causes_and_outcomes_of_sepsis_in_southeast_Asia_a_multinational_multicentre_cross-sectional_study_NCT02157259_/3486866/1.

Of 749 patients from Vietnam, 402 (54%) had no etiology identified via extensive clinical and reference laboratory workups in the original study (Fig. 1; see also Table S1 in the supplemental material); of these, 386 (96%) had clinical materials available for additional etiological investigation and were thus included for viral metagenomic analysis in this study (Fig. 1) (3). In total, 492 samples (6 CSF samples, 92 pooled nasal and throat swabs, 384 serum samples, and 10 stool samples) from these 386 patients with sepsis of unknown etiology were included in the analysis. Due to the availability of the materials, most samples were analyzed individually ($n = 458$) or in pools of multiple samples ($n = 8$) (Fig. 2).

Sample pretreatments and NA isolation. Prior to nucleic acid (NA) isolation, 100 μl of clinical sample was treated with 2 U/μl of Turbo DNase (Ambion, Life Technology, Carlsbad, CA, USA) and 0.4 U/μl RNase I (Ambion) at 37°C for 30 min. Viral NA was then isolated from nuclease-treated materials using a QIAamp viral RNA kit (Qiagen GmbH, Hilden, Germany) and was recovered in 50 μl of elution buffer.

dsDNA synthesis and sequencing. Double-stranded DNA (dsDNA) was synthesized from isolated viral NA using a set of 96 nonribosomal random primers (17), amplified by PCR, and sequenced on an Illumina MiSeq platform (Illumina, San Diego, CA, USA) as described previously (18, 19). In brief, 10 μl of extracted viral NA was converted to dsDNA using FR26RV-Endoh primers (19), SuperScript III enzyme (Invitrogen, Carlsbad, CA, USA), RNaseOUT (Invitrogen), exo-Klenow fragment (Ambion, Life Technology, Carlsbad, CA, USA), and RNase H (Ambion). Subsequently, the synthesized dsDNA was randomly amplified using the FR20RV primer (5′-GCCGGAGCTCTGCAGATATC-3′). The random PCR product obtained was then purified with the use of Agencourt AMPure XP beads (Beckman Coulter) and was quantified with a Qubit dsDNA HS (high-sensitivity) kit (Invitrogen). Finally, 1 ng of purified product was subjected to library preparation using a Nextera XT sample preparation kit (Illumina) and was sequenced using a MiSeq reagent kit, v3 (600 cycles) (Illumina), on a MiSeq platform (Illumina).

mNGS data analysis. The mNGS data were analyzed using an in-house viral metagenomic pipeline running on a 36-node Linux cluster to identify the presence of viral sequences in the tested specimens as described previously (20). In brief, after duplicate reads and reads belonging to human or bacterial genomes were filtered out, the remaining reads were assembled *de novo*. The resulting contigs and singlet reads were then aligned against a customized viral proteome database using a BLAST (Basic Local Alignment Search Tool)-based approach. Next, the candidate viral reads were aligned against a nonre-

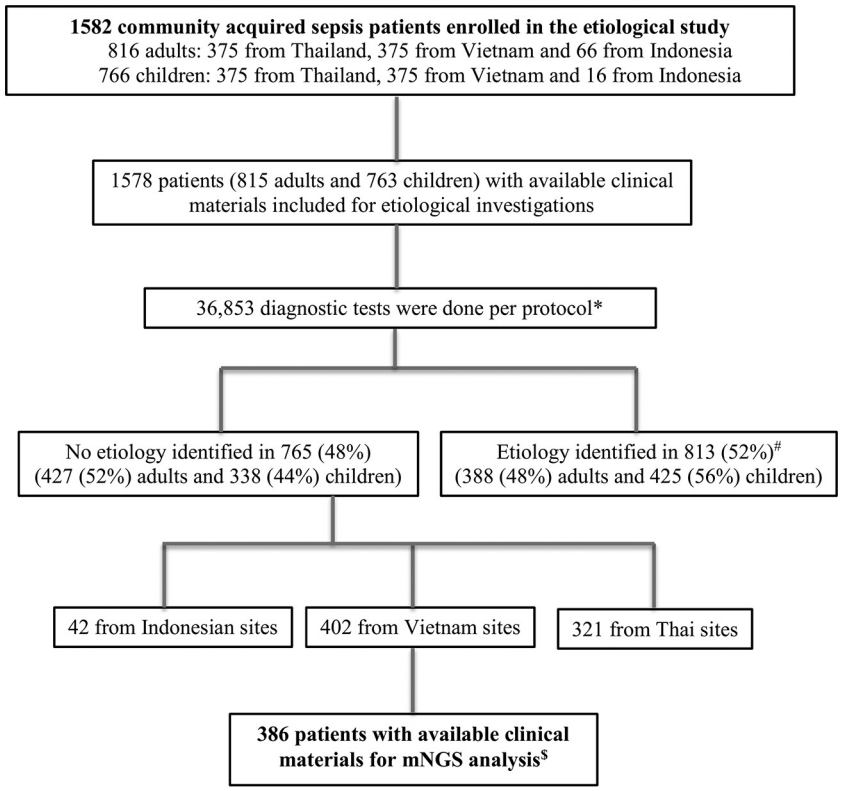


FIG 1 Flow chart showing an overview of the diagnostic output of the original study. *, see the original study (3) and Table S1 in the supplemental material for more details; #, the causative agents detected are detailed in the report of the original study (3); $, more details about the analysis of those 386 patients can be found in Fig. 2.

dundant nonvirus protein database to remove any false-positive reads (i.e., reads with expected [E] values higher than those against viral protein databases). Any virus-like sequence with an E value of $\leq 10^{-5}$ was considered a significant hit. Finally, a reference-based mapping approach was employed to assess the levels of identity and genome coverages of the corresponding viruses.

PCR confirmation of viral reads. Because of the focus of the present study, specific PCRs were used to confirm the mNGS hits for viral species that are known to be infectious to humans and for recently discovered viruses that have been reported in human tissues previously but remain of uncertain tropism. Depending on the availability of the clinical materials, virus-specific PCRs were carried out either on leftover NA after mNGS experiments or on newly extracted NA. An mNGS result was considered positive only if it was subsequently confirmed by a corresponding viral PCR analysis of original NA materials derived from the corresponding individual samples. All PCR primers and probes used were either derived from previous publications or newly designed based on the sequences generated by mNGS (see Table S2 in the supplemental material).

Phylogenetic analysis. Sequence alignment and phylogenetic tree reconstructions of the sequences obtained were carried out using ClustalW alignment and maximum likelihood methods available within Geneious 8.1.5 (Biomatters) and IQ-TREE (21), respectively.

Ethical statement. The study was reviewed and approved by the Institutional Review Boards of collaborating hospitals in Vietnam and the Oxford Tropical Research Ethics Committee (OxTREC), University of Oxford, Oxford, United Kingdom.

Accession number(s). The metagenomics data obtained in this study have been deposited in GenBank, and the accession numbers can be found via BioProject accession number PRJNA526981.

RESULTS

Demographics, clinical features, and outcomes for patients with sepsis of unknown origin. The baseline characteristics and 28-day mortality data of all patients (including the 386 patients included in the mNGS analysis) from Vietnamese sites enrolled in the original study are presented in Table 1. Retrospectively, 129 (34.4%) adult patients (including 54 of the 213 with undiagnosed cases [25%]) had SOFA (Sequential Organ Failure Assessment) scores of ≥ 2, fulfilling the diagnostic criteria presently used for sepsis in adults as defined by Sepsis-3 (22). For pediatric sepsis, no harmonized criteria similar to those for sepsis in adults have been established (23).

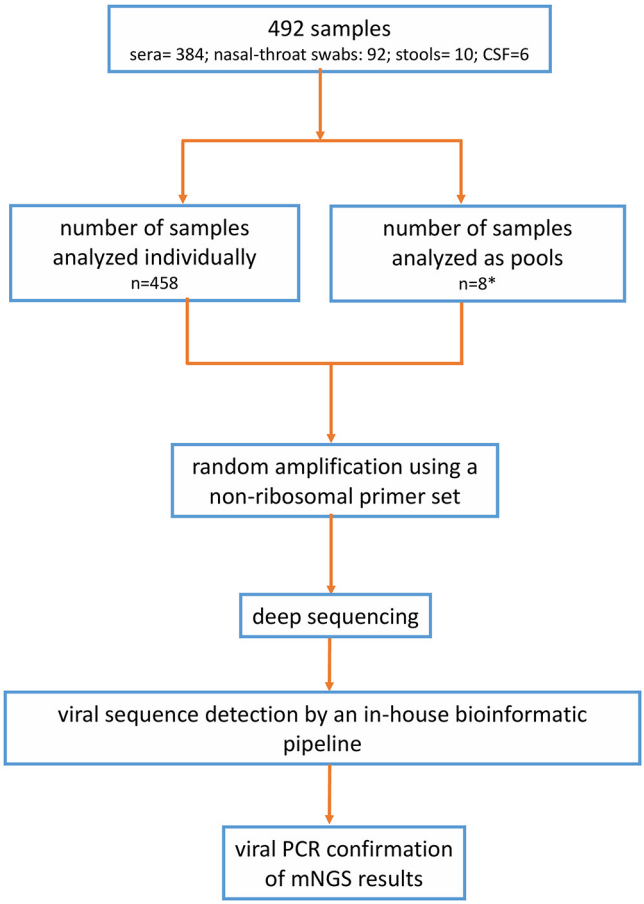


FIG 2 Flow chart showing how samples were analyzed. *, includes 4 pools of 5 samples, 3 pools of 4 samples, and 1 pool of 2 samples.

There was considerable homogeneity between the group of patients included and the group not included in the mNGS analysis (Table 1). Among the 386 patients with sepsis of unknown cause whose data were included in the mNGS analysis, the most frequent clinical entity was acute respiratory infection ($n = 158$ [41%]), followed by systemic infection ($n = 152$ [39.5%]), diarrhea ($n = 36$ [9.3%]), and central nervous system (CNS) infection ($n = 40$ [10.5%]) (Table 1) (3). Ten of these patients (8 adults and 2 children) were recorded as deceased by day 28, accounting for 2.6% of total patients.

Overview of virus-like sequences detected by mNGS. In total, 466 samples were sequenced in five MiSeq runs, generating a total of >26 million reads (median reads per sample, 432,682; range, 540 to 1,916,732) (see Fig. S1 in the supplemental material). Despite the inclusion of a nuclease digestion step prior to NA isolation, viral reads accounted for only a small proportion of total reads, ranging from 168,028 (2.5%) to 287,307 (8.4%) reads/run. Evidence of sequences related to 47 viral species belonging to 21 families was detected in 358/386 (93%) patients. The viruses detected included those known to cause human infections, those with unknown pathogenicity, and viruses that have been reported previously to be contaminants found in mNGS data sets or that have not been reported in human samples, as detailed below. Additionally, codetection of ≥2 viruses in the same samples/patients was recorded for 13 patients (see Table S3). None of the 10 fatal cases had a viral etiology identified by mNGS.

(i) Detection of viruses known to cause human infections. NA sequences of 21 viral species known to be infectious to humans were detected in 137 of 466 (29%) clinical samples from 125 of 386 (32%) individuals by viral metagenomics. The detection rate was reduced to 13.4% (52/386) of the 386 patients included in the mNGS analysis

TABLE 1 Demographic and clinical data for CA sepsis patients

Characteristic	No. (%) of patients:					
	Included in mNGS analysis[a]			Not included in mNGS analysis		
	Total (n = 386)	Adults (n = 213)	Children (n = 173)	Total (n = 363)	Adults (n = 162)	Children (n = 201)
Male gender	224 (58)	122 (57.3)	102 (59)	204 (56)	84 (41)	120 (59)
Age						
<12 mo	NA	NA	45 (26)	NA	NA	75 (37.3)
≥1 to <5 yr	NA	NA	100 (57.8)	NA	NA	106 (52.7)
≥5 to <18 yr	NA	NA	28 (16.2)	NA	NA	20 (10)
≥18 to <40 yr	NA	94 (44.1)	NA	NA	68 (42)	NA
≥40 to <60 yr	NA	67 (31.5)	NA	NA	60 (37)	NA
≥60 yr	NA	52 (24.4)	NA	NA	34 (21)	NA
Geographic location						
North Vietnam	123 (32)	68 (32)	55 (32)	127 (35)	57 (35)	70 (34)
Central Vietnam	141 (37)	79 (37)	62 (36)	108 (30)	46 (28)	62 (31)
South Vietnam	122 (32)	66 (31)	56 (32)	128 (35)	59 (37)	69 (34)
SOFA score[b]						
≤1	NA	159 (75)	NA	NA	87 (53.7)	NA
≥2	NA	54 (25)	NA	NA	75 (46.3)	NA
Clinical presentation[c]						
Respiratory infection	158 (41)	97 (45)	61 (36)	212 (58)	70 (43)	142 (71)
Diarrhea	36 (9)	25 (12)	11 (6)	15 (4)	10 (6)	5 (2)
CNS infection	40 (10.5)	8 (4)	32 (18)	42 (12)	14 (9)	28 (14)
Systemic infection	152 (39.5)	83 (39)	69 (40)	94 (26)	68 (42)	26 (13)
28-day mortality						
Yes	10 (2.6)	8 (3.7)	2 (1)	16 (4)	9 (5)	7 (3)
No	373 (96.6)	203 (95.3)	170 (98)	337 (93)	149 (92)	188 (94)
Unknown	3 (<1)	2 (1)	1 (<1)	10 (3)	4 (3)	6 (3)

[a]NA, not applicable.
[b]Available for adult patients only.
[c]Defined on the basis of major clinical symptoms. Acute respiratory infection was defined as the manifestation of at least one respiratory symptom for no longer than 14 days. Acute diarrhea was defined as diarrhea for no longer than 14 days. Acute CNS infection was defined as the manifestation of CNS symptoms for no longer than 14 days or the presence of signs of CNS infection on admission. Systemic infection was defined as the absence of acute respiratory infection, acute diarrhea, and acute CNS infection.

after specific PCR confirmation. There was a significant difference in the number of viral reads generated by mNGS between the groups of samples that were subsequently found to be PCR positive or negative (see Fig. S2 in the supplemental material), while the total numbers of reads obtained were similar for the two groups (median [range], 493,794 [11,076 to 1,203,206] versus 461,486 [16,470 to 1,770,372]) (P = 0.58). The number of reads per sample in the group of samples in which viruses were found by mNGS and subsequently confirmed by PCR was significantly higher than that in the group in which no virus was found (median [range], 493,794 [11,076 to 1,203,206] versus 365,974 [540 to 1,916,732]) (P = 0.004), suggesting that the diagnostic yield of mNGS is dependent on the sequencing depth (i.e., the number of reads generated per sample).

Of the viruses detected, enterovirus (EV) was the most common (14/386 [3.6%]), followed by hepatitis B virus (HBV) (9/386 [2.3%]), cytomegalovirus (CMV) (9/386 [2.3%]), human rhinovirus (HRV) (5/386 [1.3%]), Epstein-Barr virus (EBV) (5/386 [1.3%]), and rotavirus (3/386 [0.7%]) (Fig. 3). Detailed information about the numbers of viral reads and genome coverage is summarized in Table S7 in the supplemental material.

(ii) Detection of sequences related to viruses with unknown pathogenicity. Sequences related to four recently discovered viruses (gemycircularviruses, WU polyomavirus, human pegivirus 2 [HPgV-2], and cyclovirus-VN) whose pathogenicity or tropism remains unknown, but whose genetic materials have been reported in human samples previously, were identified by mNGS in 3.4% of the samples from the 386

FIG 3 Bar chart showing the numbers of viruses known to be infectious to humans or previously reported in human tissues that were detected by mNGS, followed by PCR confirmation testing.

patients included in the mNGS analysis. After specific PCR testing, the confirmed proportion of positive patients was reduced to 2.1% (5/386 [1.3%] had gemycircularvirus, and 1/386 [0.26%] each had WU polyomavirus, HPgV2, or cyclovirus-VN) (Fig. 3). Additionally, anellovirus-like sequences were found in the majority of the samples tested (362/466 [77%]), while sequences related to human pegivirus 1 and human papillomaviruses were found in 4/466 (<1%) and 1/466 (<1%) samples, respectively.

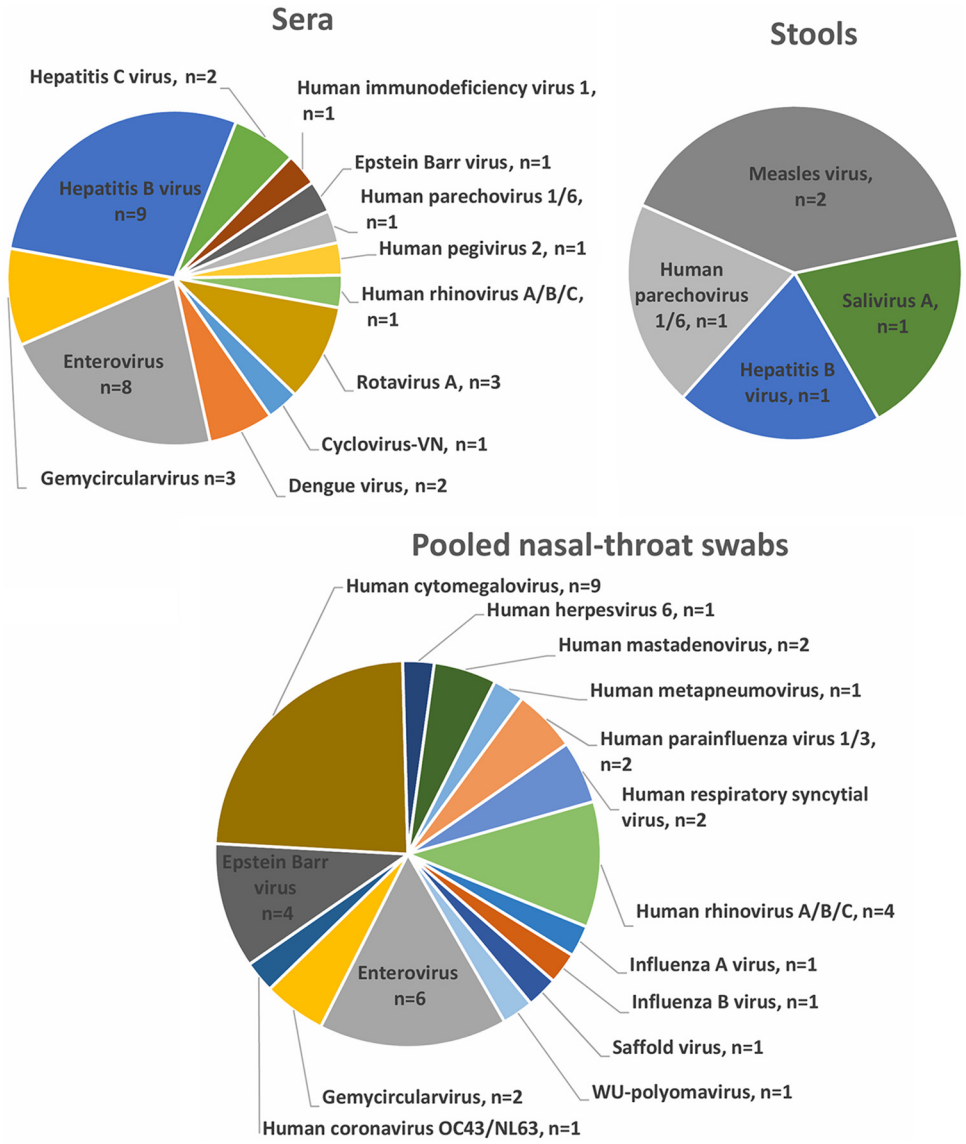


FIG 4 Numbers of viruses detected by mNGS, and then confirmed by virus-specific PCR, in clinical samples of different types.

Because these viruses are common nonpathogenic infectious agents, they were not subjected to subsequent PCR confirmation testing.

(iii) Detection of sequences related to contaminants and/or viruses not previously reported in human samples. Sequences related to common contaminants of mNGS data sets (including a parvovirus-like hybrid virus [24] and Kadipiro virus [25]) were detected in 96 and 5 samples, respectively (see Table S4 in the supplemental material). Additionally, sequences related to numerous viruses that have not been reported in human tissues previously were also found (Table S4). Here we focus our analysis on viruses that have been reported in human tissues.

Viral detection by mNGS followed by PCR confirmation testing in different sample types. The detection rates for human viruses or viruses reported in human tissues were 8% (32/384) for sera, 41% (38/92) for nasal-throat swabs, and 50% (5/10) for stool samples, while all 6 CSF samples available from 40 patients presenting with CNS infection were negative. More viruses were found in pooled nasal-throat swabs than in samples of other types (Fig. 4).

In the sera tested, 12 different viral species were detected, including the well-established human pathogens HBV ($n = 9$), EV ($n = 8$), rotavirus A ($n = 3$), dengue virus (DENV) ($n = 2$), hepatitis C virus (HCV) ($n = 2$), human parechovirus ($n = 1$), HRV ($n = 1$), and human immunodeficiency virus (HIV) ($n = 1$) (Fig. 4).

Viral detection in different patient groups and clinical entities by mNGS followed by PCR confirmation testing. The frequencies of different viral species detected in different clinical entities and patient groups are shown in Fig. 5 and Fig. S3 in the supplemental material.

Regardless of the clinical sample type, the highest proportion of distinct viral infections was recorded in patients presenting with CNS infections (15/40 [37.5%]), followed by patients with respiratory infections (37/158 [23%]) and patients with systemic infections (19/152 [12.5%]). Of the 54 adults with a SOFA score of ≥2, 6 had a virus identified (from 2 samples with measles virus or HBV and 1 each with dengue virus, rotavirus A, gemycircularvirus, salivirus A, or EBV) (see Table S5 in the supplemental material).

Among the patients presenting with CNS infections, picornaviruses were the most common viruses detected (see Table S6 in the supplemental material); these included enterovirus, accounting for 7 of 15 (47%) viruses detected (6 in sera and 1 in a pooled nasal-throat swab), and HRV, detected in a serum sample. Rotavirus, a well-known cause of diarrhea, was detected in the blood of three diarrhea patients (two children and one adult).

In terms of age groups, EV and other respiratory viruses (e.g., respiratory syncytial virus [RSV] and HRV) were detected more frequently in children than in adults (Fig. 5). In contrast, blood-borne viruses (HIV, HCV, and HBV) were found more often in adults than in children (Fig. 5). Parechovirus, an established cause of pediatric infections, was detected in one adult presenting with a systemic infection.

Genetic characterization of EV and HBV. Excluding anellovirus-related sequences, mNGS generated sufficient sequence data for informative genetic characterization and phylogenetic inference of EV and HBV in 14 samples, including seven complete viral capsid protein 1 (VP1) sequences of enterovirus and seven complete HBV genomes. Phylogenetically, all seven EVs were classified into six different serotypes of enteroviruses A and B (echovirus 3, echovirus 6, echovirus 9, echovirus 16, coxsackievirus A2, and coxsackievirus A6), while the HBV strains belonged to genotypes B and C (see Fig. S4 and S5 in the supplemental material), supporting reports about circulating enterovirus serotypes and HBV genotypes in Vietnam (26–28).

For other viruses, due to the small numbers of genomic sequences recovered (two for DENV, two for gemycircularvirus, and one each for RSV, influenza B virus, HCV, measles virus, WU polyomavirus, and cyclovirus-VN), similar phylogenetic inference was deemed uninformative.

DISCUSSION

We present the results of mNGS for exploration of the human virome in 386 patients presenting with CA sepsis of unknown cause who were enrolled in a multicenter observational study across Vietnam from 2013 to 2015. We identified 21 viral species known to be infectious to humans in 52 (13.4%) of 386 patients presenting with CA sepsis of unknown cause. The study, however, cannot directly impute sepsis causation involving the viruses identified. More specifically, on several occasions, viral detection in nonsterile materials, such as respiratory samples (including EBV and CMV) and stool samples, may simply reflect the carriage of such viruses in those bodily compartments rather than a clinical association. Similarly, viral detection (e.g., enterovirus) in the blood of patients with asymptomatic infections has been reported previously (29). Additionally, the detection of blood-borne viruses, such as HBV, HIV, and HCV, in serum samples might represent underlying diseases and not the causative pathogens leading to the hospital admission, although the detection of HIV RNA in a serum sample of a patient presenting with systemic infection may suggest an acute HIV infection. However, together with the clinical and epidemiologic data, the results present a provocative

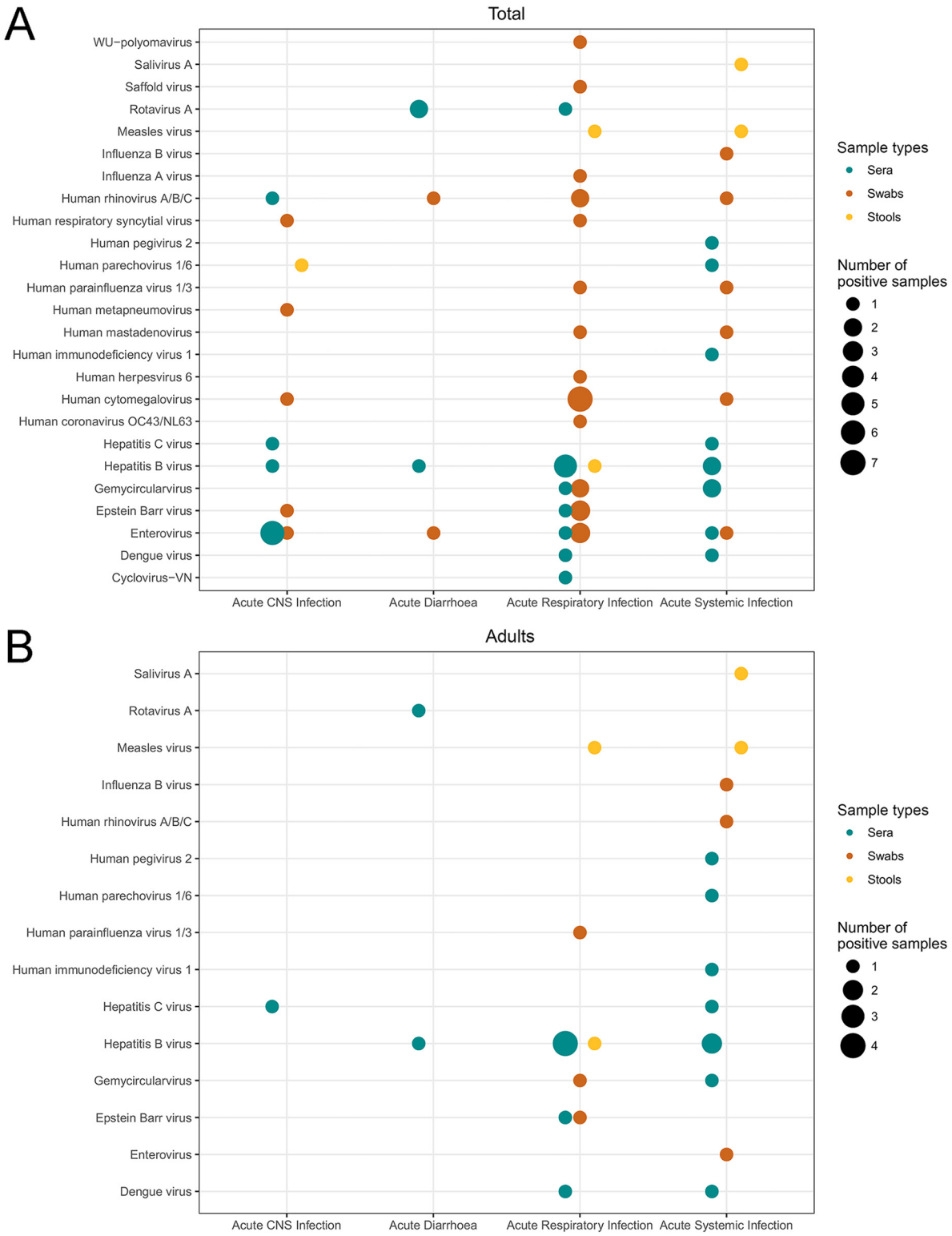


FIG 5 Numbers of viruses detected by mNGS, and then confirmed by virus-specific PCR, in different patient groups and clinical entities. Symbols are color-coded by sample type. (A) All patients included in the mNGS analysis; (B) adults; (C) children.

argument for a wide range of viral pathogens that might be associated with CA sepsis in Vietnam.

Epidemiologically, our results support previous findings regarding the frequent detection of common viruses in corresponding clinical entities and age groups. For

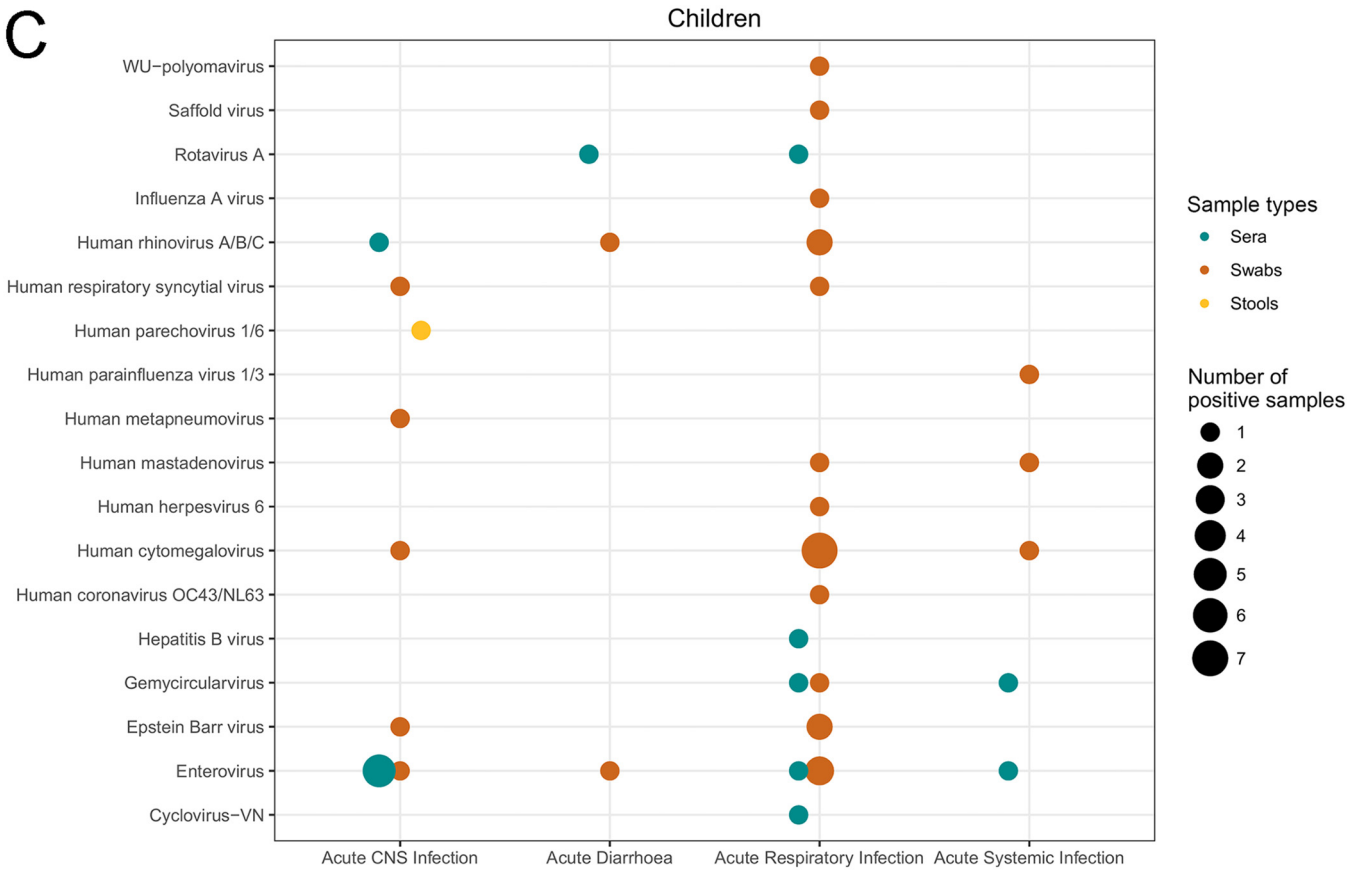


FIG 5 (Continued)

example, we found rotavirus only in patients with acute diarrhea and RSV and viruses of the *Picornaviridae* family (HRV and EV) mostly in children. Additionally, we detected parechovirus in the blood of an adult presenting with acute systemic infection. Parechoviruses are a well-known cause of disease in children, ranging from acute gastrointestinal/respiratory infections to meningitis, but have increasingly been reported to cause infections in adults (30).

Nonpolio enteroviruses, such as EV-A71 and EV-D68, have become serious global threats. In fact, EV-A71 has overwhelmed countries of the Asia-Pacific region (including Vietnam) with large outbreaks of severe hand-foot-and-mouth disease since 1997 (31, 32). Recently, EV-D68 has emerged and caused large outbreaks of respiratory infections in the United States; this virus is epidemiologically linked with acute flaccid myelitis (33). The data presented here, together with the results of the original report (3), expand our knowledge about the clinical burden posed by nonpolio enteroviruses (HRV and particularly diverse EV serotypes) and parechoviruses in Vietnam.

mNGS detected several recently discovered viruses (Saffold virus, salivirus A, WU polyomavirus, gemycircularvirus, and HPgV-2), representing their first detection in Vietnam and adding to the growing literature about the geographic distribution of these newly identified viruses. Salivirus A has been linked to gastrointestinal infection, and Saffold virus has been reported in gastrointestinal and respiratory infection patients (34–37). Saffold virus has also been reported to be associated with myocarditis and aseptic meningitis (38, 39). Additionally, using a mouse model, studies have shown the neurotropic potential of Saffold virus (39–41). The pathogenicity of WU polyomavirus, gemycircularvirus, and HPgV-2 remains unresolved. Likewise, it is imperative to conduct follow-up studies to determine whether the detected sequences that are related to viruses not previously reported in human tissues are derived from other sources and whether the respective viruses are infectious to humans.

The results of the present investigation also emphasize the utility of serum samples for assessing the etiology of sepsis. Indeed, viruses of the families *Picornaviridae* (enterovirus, rhinovirus, and parechovirus), *Flaviviridae* (DENV), and *Caliciviridae* (rotavirus) were detected by mNGS in the sera included in this study. Notably, as per the design of the original etiological study, sera were not tested for these viruses by PCR (3). Likewise, while it remains unknown why the original study failed to detect common causes of respiratory/enteric infections (influenza A virus, influenza B virus, EV, etc.) in pooled nasal swabs by multiplex PCR assays (3), a slightly lower sensitivity of the multiplex PCR assays used than that of the respective monoplex PCR assays has been reported elsewhere (42).

Virus detection by mNGS is based on the detection of matching viral reads regardless of their number or the resulting genome coverage. While few metagenomic studies published to date have reported the use of specific PCR to verify metagenomic results subsequently, the failure of virus-specific PCR to confirm the original mNGS detections for many patients in the present study may be a consequence of cross talk (bleed-through) contamination occurring as part of the sequencing procedure, a well-documented phenomenon (10, 43, 44). An alternative explanation is the low sensitivity, likely attributed to nucleotide mismatches, of some of the PCR primers used to confirm infection.

The absence of human viral pathogens in 87% of 386 patients may be attributed to the low sensitivity of our mNGS approach, especially in cases where the number of reads obtained was supposedly insufficient (Fig. S1 in the supplemental material), as suggested by the difference in the number of reads obtained between the groups of samples with and without a virus identified. Clearly, future research should address the question of what level of sequencing depth mNGS-based approaches need to achieve in order to reach the required sensitivity while maintaining cost-effectiveness. It is equally important to identify the factors (e.g., sample types and library preparation/sequencing methods) that may affect sequencing depth (i.e., the number of reads obtained) and assay sensitivity. Additional possibilities include the presence of the sepsis pathogen in nonanalyzed tissues, the presence of nonviral pathogens (e.g., bacteria and parasites) in tested specimens, and/or the inclusion of patients with no infection (e.g., those with conditions caused by toxicity whose clinical presentations mimic infections) in the study.

In summary, we report the application of mNGS for patients presenting with CA sepsis of unknown etiology. Our results highlight challenges in identifying possible viral culprits in patients with CA sepsis and show that diverse viral agents might be responsible for such devastating conditions in tropical settings such as Vietnam. Therefore, rigorous testing for a wide range of viral pathogens in samples from different body compartments collected early after symptom onset, when viral loads are usually highest, is likely to have the greatest yield. Under these circumstances, mNGS is a promising approach because of its capacity to simultaneously detect and genetically characterize viral pathogens in patient samples without the need for prior knowledge of genomic information about the targeted pathogens, thus enhancing the ability to identify infectious etiologies of sepsis and facilitating optimal targeted management.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/JCM.00386-19.

SUPPLEMENTAL FILE 1, PDF file, 0.9 MB.

ACKNOWLEDGMENTS

This study was funded by the National Institute of Allergy and Infectious Diseases, National Institutes of Health, USA (award HHSN261200800001E), and the Wellcome Trust of Great Britain (awards 106680/B/14/Z [to the Oxford University Clinical Research Unit in Vietnam] and 204904/Z/16/Z [to L.V.T.]). X.D. and E.D. were supported by the

Blood Systems Research Institute and the National Heart, Lung, and Blood Institute (grant R01 HL105770).

We thank Le Kim Thanh for logistical support. We are indebted to the patients for their participation in this study.

The Southeast Asia Infectious Disease Clinical Research Network included Pratiwi Sudarmono (Cipto Mangunkusumo Hospital, Jakarta, Indonesia); Abu Tholib Aman (Sardjito Hospital, Yogyakarta, Indonesia); Mansyur Arif (Wahidin Soedirohusodo Hospital, Makassar, Indonesia); Armaji Kamaludi Syarif, Herman Kosasih, and Muhammad Karyana (National Institute of Health Research and Development [NIHRD], Jakarta, Indonesia); Tawee Chotpitayasunondh and Warunee Punpanich Vandepitte (Queen Sirikit National Institute of Child Health, Bangkok, Thailand); Adiratha Boonyasiri, Keswadee Lapphra, Kulkanya Chokephaibulkit, Pinyo Rattanaumpawan, and Visanu Thamlikitkul (Siriraj Hospital, Bangkok, Thailand); Achara Laongnualpanich (Chiang Rai Prachanukroh Hospital, Chiang Rai, Thailand); Prapit Teparrakkul and Pramot Srisamang (Sappasithiprasong Hospital, Ubon Ratchathani, Thailand); Phan Huu Phuc and Le Thanh Hai (National Hospital of Paediatrics, Hanoi, Vietnam); Nguyen Van Kinh (National Hospital of Tropical Diseases, Hanoi, Vietnam); Bui Duc Phu (Hue Central Hospital, Hue, Vietnam); Nguyen Thanh Hung and Tang Chi Thuong (Children's Hospital 1, Ho Chi Minh City, Vietnam); Ha Manh Tuan (Children's Hospital 2, Ho Chi Minh City, Vietnam); Lam Minh Yen and Nguyen Van Vinh Chau (Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam); Direk Limmathurotsakul, Janjira Thaipadungpanit, Stuart Blacksell, and Nicholas Day (Mahidol-Oxford Tropical Medicine Research Unit [MORU], Bangkok, Thailand); Claire Ling (Shoklo Malaria Research Unit, Mae Sot, Tak, Thailand); Guy Thwaites, Heiman Wertheim, Le Van Tan, Motiur Rahman, and H. Rogier van Doorn (Oxford University Clinical Research Unit [OUCRU], Ho Chi Minh City, Vietnam); and Chuen-Yen Lau (National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA).

REFERENCES

1. Reinhart K, Daniels R, Kissoon N, Machado FR, Schachter RD, Finfer S. 2017. Recognizing sepsis as a global health priority—a WHO resolution. N Engl J Med 377:414–417. https://doi.org/10.1056/NEJMp1707170.
2. Schultz MJ, Dunser MW, Dondorp AM, Adhikari NKJ, Iyer S, Kwizera A, Lubell Y, Papali A, Pisani L, Riviello BD, Angus DC, Azevedo LC, Baker T, Diaz JV, Festic E, Haniffa R, Jawa R, Jacob ST, Kissoon N, Lodha R, Martin-Loeches I, Lundeg G, Misango D, Mer M, Mohanty S, Murthy S, Musa N, Nakibuuka J, Serpa Neto A, Nguyen Thi Hoang M, Nguyen Thien B, Pattnaik R, Phua J, Preller J, Povoa P, Ranjit S, Talmor D, Thevanayagam J, Thwaites CL; Global Intensive Care Working Group of the European Society of Intensive Care Medicine. 2017. Current challenges in the management of sepsis in ICUs in resource-poor settings and suggestions for the future. Intensive Care Med 43:612–624. https://doi.org/10.1007/s00134-017-4750-z.
3. Southeast Asia Infectious Disease Clinical Research Network. 2017. Causes and outcomes of sepsis in Southeast Asia: a multinational multicentre cross-sectional study. Lancet Glob Health 5:e157–e167. https://doi.org/10.1016/S2214-109X(17)30007-4.
4. Allen T, Murray KA, Zambrana-Torrelio C, Morse SS, Rondinini C, Di Marco M, Breit N, Olival KJ, Daszak P. 2017. Global hotspots and correlates of emerging zoonotic diseases. Nat Commun 8:1124. https://doi.org/10.1038/s41467-017-00923-8.
5. Baud D, Gubler DJ, Schaub B, Lanteri MC, Musso D. 2017. An update on Zika virus infection. Lancet 390:2099–2109. https://doi.org/10.1016/S0140-6736(17)31450-2.
6. Nakamura S, Yang CS, Sakon N, Ueda M, Tougan T, Yamashita A, Goto N, Takahashi K, Yasunaga T, Ikuta K, Mizutani T, Okamoto Y, Tagami M, Morita R, Maeda N, Kawai J, Hayashizaki Y, Nagai Y, Horii T, Iida T, Nakaya T. 2009. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PLoS One 4:e4219. https://doi.org/10.1371/journal.pone.0004219.
7. Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, Conlan S, Quan PL, Hui J, Marshall J, Simons JF, Egholm M, Paddock CD, Shieh WJ, Goldsmith CS, Zaki SR, Catton M, Lipkin WI. 2008. A new arenavirus in a cluster of fatal transplant-associated diseases. N Engl J Med 358:991–998. https://doi.org/10.1056/NEJMoa073785.
8. Tan LV, van Doorn HR, Nghia HDT, Chau TTH, Tu LTP, de Vries M, Canuti M, Deijs M, Jebbink MF, Baker S, Bryant JE, Tham NT, B'Krong NTTC, Boni MF, Loi TQ, Phuong LT, Verhoeven JTP, Crusat M, Jeeninga RE, Schultsz C, Chau NVV, Hien TT, van der Hoek L, Farrar J, de Jong MD. 2013. Identification of a new cyclovirus in cerebrospinal fluid of patients with acute central nervous system infections. MBio 4:e00231-13. https://doi.org/10.1128/mBio.00231-13.
9. Towner JS, Sealy TK, Khristova ML, Albariño CG, Conlan S, Reeder SA, Quan PL, Lipkin WI, Downing R, Tappero JW, Okware S, Lutwama J, Bakamutumaho B, Kayiwa J, Comer JA, Rollin PE, Ksiazek TG, Nichol ST. 2008. Newly discovered Ebola virus associated with hemorrhagic fever outbreak in Uganda. PLoS Pathog 4:e1000212. https://doi.org/10.1371/journal.ppat.1000212.
10. Yang J, Yang F, Ren L, Xiong Z, Wu Z, Dong J, Sun L, Zhang T, Hu Y, Du J, Wang J, Jin Q. 2011. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. J Clin Microbiol 49:3463–3469. https://doi.org/10.1128/JCM.00273-11.
11. Edridge AWD, Deijs M, Namazzi R, Cristella C, Jebbink MF, Maurer I, Kootstra NA, Buluma LR, van Woensel JBM, de Jong MD, Idro R, Boele van Hensbroek M, van der Hoek L. 2019. Novel orthobunyavirus identified in the cerebrospinal fluid of a Ugandan child with severe encephalopathy. Clin Infect Dis 68:139–142. https://doi.org/10.1093/cid/ciy486.
12. Noguera-Julian M, González-Beiras C, Parera M, Ubals M, Kapa A, Paredes R, Mitjà O. 2019. Etiological characterization of the cutaneous ulcer syndrome in Papua New Guinea using shotgun metagenomics. Clin Infect Dis 68:482–489. https://doi.org/10.1093/cid/ciy502.
13. Piantadosi A, Kanjilal S, Ganesh V, Khanna A, Hyle EP, Rosand J, Bold T, Metsky HC, Lemieux J, Leone MJ, Freimark L, Matranga CB, Adams G, McGrath G, Zamirpour S, Telford S, Rosenberg E, Cho T, Frosch MP, Goldberg MB, Mukerji SS, Sabeti PC. 2018. Rapid detection of Powassan virus in a patient with encephalitis by metagenomic sequencing. Clin Infect Dis 66:789–792. https://doi.org/10.1093/cid/cix792.

14. Dennis TPW, Flynn PJ, de Souza WM, Singer JB, Moreau CS, Wilson SJ, Gifford RJ. 2018. Insights into circovirus host range from the genomic fossil record. J Virol 92:e00145-18. https://doi.org/10.1128/JVI.00145-18.
15. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, Sevransky JE, Sprung CL, Douglas IS, Jaeschke R, Osborn TM, Nunnally ME, Townsend SR, Reinhart K, Kleinpell RM, Angus DC, Deutschman CS, Machado FR, Rubenfeld GD, Webb SA, Beale RJ, Vincent JL, Moreno R; Surviving Sepsis Campaign Guidelines Committee including the Pediatric Subgroup. 2013. Surviving Sepsis Campaign : international guidelines for management of severe sepsis and septic shock : 2012. Crit Care Med 41:580–637. https://doi.org/10.1097/CCM.0b013e31827e83af.
16. Goldstein B, Giroir B, Randolph A; International Consensus conference on Pediatric Sepsis. 2005. International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics. Pediatr Crit Care Med 6:2–8. https://doi.org/10.1097/01.PCC.0000149131.72248.E6.
17. Endoh D, Mizutani T, Kirisawa R, Maki Y, Saito H, Kon Y, Morikawa S, Hayashi M. 2005. Species-independent detection of RNA virus by representational difference analysis using non-ribosomal hexanucleotides for reverse transcription. Nucleic Acids Res 33:e65. https://doi.org/10.1093/nar/gni064.
18. Tan Le V, van Doorn HR, van der Hoek L, Minh Hien V, Jebbink MF, Quang Ha D, Farrar J, Van Vinh Chau N, de Jong MD. 2011. Random PCR and ultracentrifugation increases sensitivity and throughput of VIDISCA for screening of pathogens in clinical specimens. J Infect Dev Ctries 5:142–148. https://doi.org/10.3855/jidc.1087.
19. Froussard P. 1992. A random-PCR method (rPCR) to construct whole cDNA library from low amounts of RNA. Nucleic Acids Res 20:2900. https://doi.org/10.1093/nar/20.11.2900.
20. Aiemjoy K, Altan E, Aragie S, Fry DM, Phan TG, Deng X, Chanyalew M, Tadesse Z, Callahan EK, Delwart E, Keenan JD. 2019. Viral species richness and composition in young children with loose or watery stool in Ethiopia. BMC Infect Dis 19:53. https://doi.org/10.1186/s12879-019-3674-3.
21. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300.
22. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Coopersmith CM, Hotchkiss RS, Levy MM, Marshall JC, Martin GS, Opal SM, Rubenfeld GD, van der Poll T, Vincent JL, Angus DC. 2016. The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA 315:801–810. https://doi.org/10.1001/jama.2016.0287.
23. Obonyo NG, Schlapbach LJ, Fraser JF. 2019. Sepsis: changing definitions, unchanging treatment. Front Pediatr 6:425. https://doi.org/10.3389/fped.2018.00425.
24. Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, Aronsohn A, Hackett J, Delwart EL, Chiu CY. 2013. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. J Virol 87:11966–11977. https://doi.org/10.1128/JVI.02323-13.
25. Ngoi CN, Siqueira J, Li L, Deng X, Mugo P, Graham SM, Price MA. 2017. Corrigendum: The plasma virome of febrile adult Kenyans shows frequent parvovirus B19 infections and a novel arbovirus (Kadipiro virus). J Gen Virol 98:517. https://doi.org/10.1099/jgv.0.000762.
26. B'Krong NTTC, Minh NNQ, Qui PT, Chau TTH, Nghia HDT, Do LAH, Nhung NN, Van Vinh Chau N, Thwaites G, Van Tan L, van Doorn HR, Thanh TT. 2018. Enterovirus serotypes in patients with central nervous system and respiratory infections in Viet Nam 1997–2010. Virol J 15:69. https://doi.org/10.1186/s12985-018-0980-0.
27. Nguyen TA, Le NTN, Van Hoang MT, Nguyen TTH, Tran TT, Vu TTH, Nguyen THN, Lam AN, Tran T, Lan P, Le NTN, Nguyen TH, Truong HK, Ha MT, Ho LV, Nguyen TN, Do CV, Phan TQ, Wills B, Sabanathan S, Nguyen VVC, Thwaites L, Van Doorn HR, Thwaites G, Rabaa MA, Le VT. 2018. Emerging coxsackievirus A6 causing hand, foot and mouth disease, Vietnam. Emerg Infect Dis 24:17–19. https://doi.org/10.1016/j.ijid.2016.02.078.
28. Bui TTT, Tran TT, Nghiem MN, Rahman P, Tran TTT, Dinh MNH, Le MH, Nguyen VVC, Thwaites G, Rahman M. 2017. Molecular characterization of hepatitis B virus in Vietnam. BMC Infect Dis 17:601. https://doi.org/10.1186/s12879-017-2697-x.
29. Honkanen H, Oikarinen S, Pakkanen O, Ruokoranta T, Pulkki MM, Laitinen OH, Tauriainen S, Korpela S, Lappalainen M, Vuorinen T, Haapala AM, Veijola R, Simell O, Ilonen J, Knip M, Hyöty H. 2013. Human enterovirus 71 strains in the background population and in hospital patients in Finland. J Clin Virol 56:348–353. https://doi.org/10.1016/j.jcv.2012.11.018.
30. Kong KL, Lau JSY, Goh SM, Wilson HL, Catton M, Korman TM. 2017. Myocarditis caused by human parechovirus in adult. Emerg Infect Dis 23:1571–1573. https://doi.org/10.3201/eid2309.161256.
31. Ooi MH, Wong SC, Lewthwaite P, Cardosa MJ, Solomon T. 2010. Clinical features, diagnosis, and management of enterovirus 71. Lancet Neurol 9:1097–1105. https://doi.org/10.1016/S1474-4422(10)70209-X.
32. Nhan LNT, Hong NTT, Nhu LNT, Nguyet LA, Ny NTH, Thanh TT, Han DDK, Van HMT, Thwaites CL, Hien TT, Qui PT, Quang PV, Minh NNQ, van Doorn HR, Khanh TH, Chau NVV, Thwaites G, Hung NT, Tan LV. 2018. Severe enterovirus A71 associated hand, foot and mouth disease, Vietnam, 2018 : preliminary report of an impending outbreak. Euro Surveill 23. https://doi.org/10.2807/1560-7917.ES.2018.23.46.1800590.
33. Messacar K, Asturias EJ, Hixon AM, Van Leer-Buter C, Niesters HGM, Tyler KL, Abzug MJ, Dominguez SR. 2018. Enterovirus D68 and acute flaccid myelitis—evaluating the evidence for causality. Lancet Infect Dis 18:e239–e247. https://doi.org/10.1016/S1473-3099(18)30094-X.
34. Tan SZK, Tan MZY, Prabakaran M. 2017. Saffold virus, an emerging human cardiovirus. Rev Med Virol 27. https://doi.org/10.1002/rmv.1908.
35. Drexler JF, de Souza Luna LK, Stöcker A, Silva Almeida P, Medrado Ribeiro TC, Petersen N, Herzog P, Pedroso C, Huppertz HI, da Costa Ribeiro H, Jr, Baumgarte S, Drosten C. 2008. Circulation of 3 lineages of a novel Saffold cardiovirus in humans. Emerg Infect Dis 14:1398–1405. https://doi.org/10.3201/eid1409.080570.
36. Nielsen ACY, Böttiger B, Banner J, Hoffmann T, Nielsen LP. 2012. Serious invasive Saffold virus infections in children, 2009. Emerg Infect Dis 18:7–12. https://doi.org/10.3201/eid1801.110725.
37. Reuter G, Pankovics P, Boros Á. 2017. Saliviruses—the first knowledge about a newly discovered human picornavirus. Rev Med Virol 27. https://doi.org/10.1002/rmv.1904.
38. Nielsen TS, Nielsen AY, Banner J, Hansen J, Baandrup U, Nielsen LP. 2016. Saffold virus infection associated with human myocarditis. J Clin Virol 74:78–81. https://doi.org/10.1016/j.jcv.2015.11.028.
39. Himeda T, Hosomi T, Asif N, Shimizu H, Okuwa T, Muraki Y, Ohara Y. 2011. The preparation of an infectious full-length cDNA clone of Saffold virus. Virol J 8:110. https://doi.org/10.1186/1743-422X-8-110.
40. Hertzler S, Liang Z, Treso B, Lipton HL. 2011. Adaptation of Saffold virus 2 for high-titer growth in mammalian cells. J Virol 85:7411–7418. https://doi.org/10.1128/JVI.00265-11.
41. Tan SZK, Chua KB, Xu Y, Prabakaran M. 2016. The pathogenesis of Saffold virus in AG129 mice and the effects of its truncated L protein in the central nervous system. Viruses 8:E24. https://doi.org/10.3390/v8020024.
42. Jansen RR, Schinkel J, Koekkoek S, Pajkrt D, Beld M, de Jong MD, Molenkamp R. 2011. Development and evaluation of a four-tube real time multiplex PCR assay covering fourteen respiratory viruses, and comparison to its corresponding single target counterparts. J Clin Virol 51:179–185. https://doi.org/10.1016/j.jcv.2011.04.010.
43. Wilson MR, Fedewa G, Stenglein MD, Olejnik J, Rennick LJ, Nambulli S, Feldmann F, Duprex WP, Connor JH, Mühlberger E, DeRisi JL. 2016. Multiplexed metagenomic deep sequencing to analyze the composition of high-priority pathogen reagents. mSystems 1:e00058-16. https://doi.org/10.1128/mSystems.00058-16.
44. Mitra A, Skrzypczak M, Ginalski K, Rowicka M. 2015. Strategies for achieving high sequencing accuracy for low diversity samples and avoiding sample bleeding using Illumina platform. PLoS One 10:e0120520. https://doi.org/10.1371/journal.pone.0120520.

Detection and Characterization of Human Pegivirus 2, Vietnam

Nguyen To Anh, Nguyen Thi Thu Hong, Le Nguyen Truc Nhu, Tran Tan Thanh, Catherine Anscombe, Le Ngoc Chau, Tran Thi Thanh Thanh, Chuen-Yen Lau, Direk Limmathurotsakul, Nguyen Van Vinh Chau, H. Rogier van Doorn, Xutao Deng, Motiur Rahman, Eric Delwart, Thuy Le, Guy Thwaites, Le Van Tan, for the Southeast Asia Infectious Disease Clinical Research Network[1]

We report human pegivirus 2 (HPgV-2) infection in Vietnam. We detected HPgV-2 in some patients with hepatitis C virus/HIV co-infection but not in patients with HIV or hepatitis A, B, or C virus infection, nor in healthy controls. HPgV-2 strains in Vietnam are phylogenetically related to global strains.

Human pegivirus 2 (HPgV-2), also known as human hepegivirus 1, is a recently discovered bloodborne flavivirus (*1*,*2*). Existing evidence suggests that HPgV-2 is more frequently detected in patients with hepatitis C virus (HCV) infection, particularly HCV and HIV co-infection, although detection rates vary between studies and patient groups. In the United States, HPgV-2 was detected in 1.2% (12/983) of patients with active HCV infections (*1*), whereas in China, the reported detection rates of HPgV-2 RNA were 0.29% (7/2440) among HCV monoinfected patients and from 3% (8/270) to 5.7% (4/70) among HCV/HIV co-infected patients (*3*,*4*). HPgV-2 RNA was detected in 10.9% (17/156) of injection drug users in the United States, and there was a strong association between HPgV-2 and other infections such as HCV and SEN virus D (*5*).

Author affiliations: Oxford University, Ho Chi Minh City, Vietnam (N.T. Anh, N.T.T. Hong, L.N.T. Nhu, T.T. Thanh, C. Anscombe, L.N. Chau, T.T.T. Thanh, H.R. van Doorn, M. Rahman, T. Le, G. Thwaites, L.V. Tan); University of Oxford, Oxford, UK (C. Anscombe, D. Limmathurotsakul, H.R. van Doorn, G. Thwaites); National Institutes of Health, Bethesda, Maryland, USA (C.-Y. Lau); Mahidol Oxford Tropical Research Unit, Bangkok, Thailand (D. Limmathurotsakul); Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam (N.V.V. Chau); Blood Systems Research Institute, San Francisco, California, USA (X. Deng, E. Delwart); University of California, San Francisco (X. Deng, E. Delwart)

DOI: https://doi.org/10.3201/eid2411.180668

Given the high burden of HCV and HIV infections worldwide and the potential clinical significance of HPgV-2, we investigated the geographic distribution and genetic diversity of this virus to help prioritize the development and implementation of appropriate intervention strategies. The studies were approved by the corresponding institutional review boards of the local hospitals in Vietnam where patients were enrolled and the Oxford Tropical Research Ethics Committee. We obtained written informed consent from each participant or from the participant's parent or legal guardian.

The Study

Patient information and clinical samples were derived from a multilocation observational study designed to evaluate the causes of community-acquired infection in Southeast Asia (*6*). We included all 493 samples (384 plasma, 92 pooled nasal and throat swabs, 10 stool, and 7 cerebrospinal fluid [CSF]) from 386 patients in Vietnam with community-acquired infection of unknown origin after extensive diagnostic workup for viral metagenomic analysis (*7*).

Analysis of metagenomic data revealed that, in 1 plasma sample, of 98,344 obtained reads, 5,342 reads were of HCV sequences, 430 of HIV sequences, and 273 of HPgV-2 sequences; we confirmed all reads by corresponding virus-specific reverse transcription PCR (RT-PCR). HPgV-2 sequence screening and HPgV-2 RT-PCR testing did not detect HPgV-2 in any of the remaining samples of the patients included in metagenomic analysis.

To explore the prevalence of HPgV-2 in HCV-infected patients in Vietnam, we used a reference-based mapping strategy to screen for HPgV-2 sequences in additional viral metagenomic datasets (Table 1). We detected HPgV-2 sequences in 5/79 HIV/HCV co-infected patients who participated in a trial evaluating the hepatic safety of raltegravir/efavirenz-based therapies in antiretroviral-naive HIV-infected subjects co-infected with HCV. We did not detect HPgV-2 sequences in 394 HCV-infected patients with clinically diagnosed hepatitis who participated in molecular epidemiologic studies of hepatitis viruses (Table 1).

We subsequently confirmed the result of this reference-mapping approach by HPgV-2 multiplex RT-PCR (*8*) testing of the extracted RNA from original samples. We

[1]Members of the Southeast Asia Infectious Disease Clinical Research Network are listed at the end of this article.

Table 1. Samples and viral metagenomic datasets used in screening for HPgV-2 and screening results, Vietnam*

Infection	No. persons	Screening approach	No. positive for HPgV-2	Enrollment period	Setting
Hepatitis C virus and HIV co-infection	79	HPgV-2–specific PCR and reference-based mapping of obtained viral metagenomics data	5	2010–2013	Hospital for Tropical Diseases, Ho Chi Minh City
HIV monoinfection	78	HPgV-2–specific PCR	0	2010–2013	Hospital for Tropical Diseases, Ho Chi Minh City
None (healthy volunteers)	80	HPgV-2–specific PCR	0	2010–2013	Hospital for Tropical Diseases, Ho Chi Minh City
Hepatitis A virus	71	HPgV-2–specific PCR	0	2012–2014	Hospital for Tropical Diseases, Ho Chi Minh City
Hepatitis B virus	103	HPgV-2–specific PCR	0	2012–2016	Hospital for Tropical Diseases, Ho Chi Minh City; Dong Thap General Hospital, Dong Thap; Khanh Hoa Provincial Hospital, Nha Trang; Dac Lac Provincial Hospital, Dac Lac; Hue National Hospital, Hue
Hepatitis C virus†	394	Reference-based mapping of obtained viral metagenomics data	0	2012–2016	Hospital for Tropical Diseases, Ho Chi Minh City; Dong Thap General Hospital, Dong Thap; Khanh Hoa Provincial Hospital, Nha Trang; Dac Lac Provincial Hospital, Dac Lac; Hue National Hospital, Hue

*HPgV-2, human pegivirus.
†Whole-genome sequences of hepatitis C virus were obtained using a viral metagenomics approach (*7*). The resulting metagenomics datasets were then subjected to a reference-based mapping approach to search for the presence of HPgV-2 sequences.

conducted multiplex RT-PCR screening for HPgV-2 RNA in plasma samples of matched controls (78 HIV-infected patients and 80 healthy volunteers) of the 79 HCV/HIV co-infected patients; we found no evidence of HPgV-2 (Table 1). In addition, we did not detect HPgV-2 RNA in any plasma samples from patients with HAV (n = 71) and HBV (n = 103) infection (Table 1).

HPgV-2 RNA was detectable for ≤18 months in 3/5 patients with HCV/HIV co-infection (Table 2). We did not detect HPgV-2 RNA in the available follow-up serum sample collected 14 days after enrollment from the patient with community-acquired infection (Table 2).

All 5 HCV/HIV co-infected patients had CD4 counts >200 cells/μL at baseline and at 6-, 12-, and 18-month follow-up (Table 2), but none received specific anti-HCV drugs, which was attributed to drug unavailability or unaffordability during the study period. During follow-up, hepatitis and splenic abnormalities were detected in 4/5 patients, which were likely attributable to HCV infection (Table 2). The patient with community-acquired infection was recorded as surviving to 28 days of follow-up (Table 2).

Using deep sequencing and a combination of overlapping PCRs and Sanger sequencing of PCR amplicons (primer sequences available upon request), we obtained 5 nearly complete genomes (coverage of >92%) and another partial genome (coverage of ≈69.1%). Pairwise comparison of HPgV-2 coding regions obtained in this study and previously reported HPgV-2 sequences showed overall sequence identities at the nucleotide level of ≥94.6% and at the amino acid level of ≥95.3% (data not shown). Phylogenetic analyses revealed a tight cluster between viruses from Vietnam and global strains (Figure). We submitted the HPgV-2 sequences we generated to GenBank (accession nos. MH194408–13).

Of the 5 HPgV-2 genome sequences we recovered, we generated 2 by deep sequencing. The results were above the proposed sequencing-depth threshold of ≥5 for sequences generated by next-generation sequencing (*9*) and sufficient for intrahost diversity investigation. One sequence we generated had mean coverage of 2,049 (range 12–9,912), with a total of 26 (10 [38%] nonsynonymous) positions carrying minor variations detected in the corresponding dataset (data not shown). For the other sequence, mean coverage was 32,531 (range 13–138,383), with a total of 37 (13 [35%] nonsynonymous) positions carrying minor variations in its dataset (data not shown).

Conclusions

We report the detection and genetic characterization of HPgV-2 in Vietnam and describe the observed demographic and clinical characteristics of patients with HPgV-2 infection. Together with reports from China, Iran, and the United States (*1–4*,*8*,*10*), our findings further emphasize the strong association between HPgV-2 and HCV, especially HCV/HIV co-infection. The absence of HPgV-2 in 394 HCV-infected patients may have been attributed to the small sample size and the fact that the reported prevalence of HIV among HCV-infected patients was ≤6.5% (*11*,*12*). Of note, HPgV-2 was detected in only 0.29% of HCV-monoinfected patients in China.

Table 2. Demographic and clinical features of 6 men with human pegivirus infection, Vietnam*

Pt no.	Pt age, y	Time point, mo	HCV RNA+	HPgV-2 RNA+	Total bilirubin, μmol/L	Direct bilirubin, μmol/L	AST, UI/L	ALT, UI/L	CD4 count, cells/μL	HIV RNA, × 10^3 copies/μL	AFP, mg/mL	FibroScan result, kPa	Symptoms
1	29	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2	47	0	Y	Y	9.8	0.7	30	24	331	120	1.7	11.8	
		6	Y	Y	4.7	1.6	81	83	518	0.07	2.3	NA	
		12	Y	Y	6.9	3.4	55	61	364	0.04	2.6	11.8	Hepatitis
		18	Y	Y	4.8	2.8	37	40	428	UND	2.14	6.1	Hepatomegaly
3	32	0	Y	Y	4.7	3.4	39	10	288	0.198	0.999	6.5	
		6	Y	Y	12.8	4.7	50	19	510	0.04	1.68	NA	
		12	Y	Y	9.5	5.3	63	25	622	UND	1.88	6.2	Liver fibrosis, hepatomegaly
		18	Y	Y	7.6	3.8	42	23	622	UND	1.53	7.2	Hepatitis
4	35	0	Y	Y	7.8	4.9	67	55	290	61.1	2.96	6.4	
		6	Y	Y	10.7	6.3	77	80	411	UND	3.1	NA	
		12	Y	Y	8.8	3.9	76	72	337	UND	4	8.5	Homogeneous hepatomegaly
		18	Y	Y	13	6.3	108	129	455	UND	4.1	8.1	Splenomegaly, liver fibrosis
5	34	0	Y	Y	4.3	2.8	33	43	291	70.2	3.67	6.1	
		6	N	Y	6.5	2.1	35	43	287	UND	3.83	NA	
		12	N	N	5.4	2.6	33	40	484	UND	4.48	4.5	
		18	N	N	6.6	2.6	73	85	546	UND	3.9	3	
6	31	0	Y	Y	4.5	2.4	52.2	36.5	295	96.8	12.7	22.8	
		6	Y	Y	17.1	12.9	64	62	579	UND	16.74	NA	
		12	Y	N	12.3	4.3	114	121	711	UND	46.3	26.3	Mild liver fibrosis, mild splenomegaly
		18	Y	N	10.6	4.9	82	89	816	UND	61.01	NA	Hepatomegaly, splenomegaly

*Age is patient's age at diagnosis; time point is the month at which follow-up visit was conducted; 0 was the baseline examination. ALT, alanine aminotransferase; AS, aspartate aminotransferase; NA, not available; Pt, patient; UND, undetectable.
†Patient 1 belongs to the community-acquired infection cohort.

Previous reports showed that HPgV-2 viremia can be transient or persistent. Likewise, in our study, HPgV-2 RNA became undetectable after 14 days in a HCV/HIV co-infected patient with community-acquired infection of unknown origin, but remained detectable in other HCV/HIV co-infected patients through up to 18 months of follow-up.

The pathogenic potential of HPgV-2 remains unknown. Its role in HCV/HIV co-infection and response to treatment warrants further research, given its low detection rates in blood donors in the United States and China (*1*,*3*) and its absence in healthy persons (this study) but close association with HCV/HIV co-infection.

In the era of sequence-based virus discovery, a key question is whether the detected genome represents live virus or a non–replication competent genome. Addressing this question would require recovery of virus in cell culture. However, our detection of minor variations across 2 HPgV-2 genomes suggests that viral replication had occurred in the infected patients. Phylogenetically, the close relatedness between HPgV-2 strains from Vietnam and global strains suggests HPgV-2 has a wide geographic distribution.

Our study has some limitations. First, we only retrospectively tested available archived samples without formal sample size estimation, which may have explained the absence of HPgV-2 in the remaining 394 HCV patients. Second, we did not employ a serologic assay to screen for HPgV-2 antibodies in patients' plasma. Third, we used only multiplex PCR with primers based on a limited number of available HPgV-2 sequences. Therefore, we may have missed genetically diverse HPgV-2 strains, and we may have underestimated the prevalence of HPgV-2 infections in Vietnam.

Our findings contribute expanded data about geographic distribution, demographics, and genetic diversity of HPgV-2. Because HCV and HIV infections are global public health issues, the extent to which HPgV-2 interacts with HCV and HIV in co-infected patients and the possible clinical consequences warrant further research.

Members of the Southeast Asia Infectious Disease Clinical Research Network: Pratiwi Sudarmono (Cipto Mangunkusumo Hospital, Jakarta, Indonesia); Abu Tholib Aman (Sardjito Hospital, Yogyakarta, Indonesia); Mansyur Arif (Wahidin Soedirohusodo Hospital, Makassar, Indonesia); Armaji Kamaludi Syarif, Herman Kosasih, and Muhammad Karyana (National Institute of Health Research and Development, Jakarta); Tawee Chotpitayasunondh and Warunee Punpanich Vandepitte (Queen Sirikit National Institute of Child Health, Bangkok, Thailand); Adiratha Boonyasiri, Keswadee Lapphra, Kulkanya Chokephaibulkit, Pinyo Rattanaumpawan, and Visanu Thamlikitkul (Siriraj Hospital, Bangkok, Thailand); Direk Limmathurotsakul, Janjira Thaipadungpanit, Stuart Blacksell, and Nicholas Day (Mahidol-Oxford Tropical

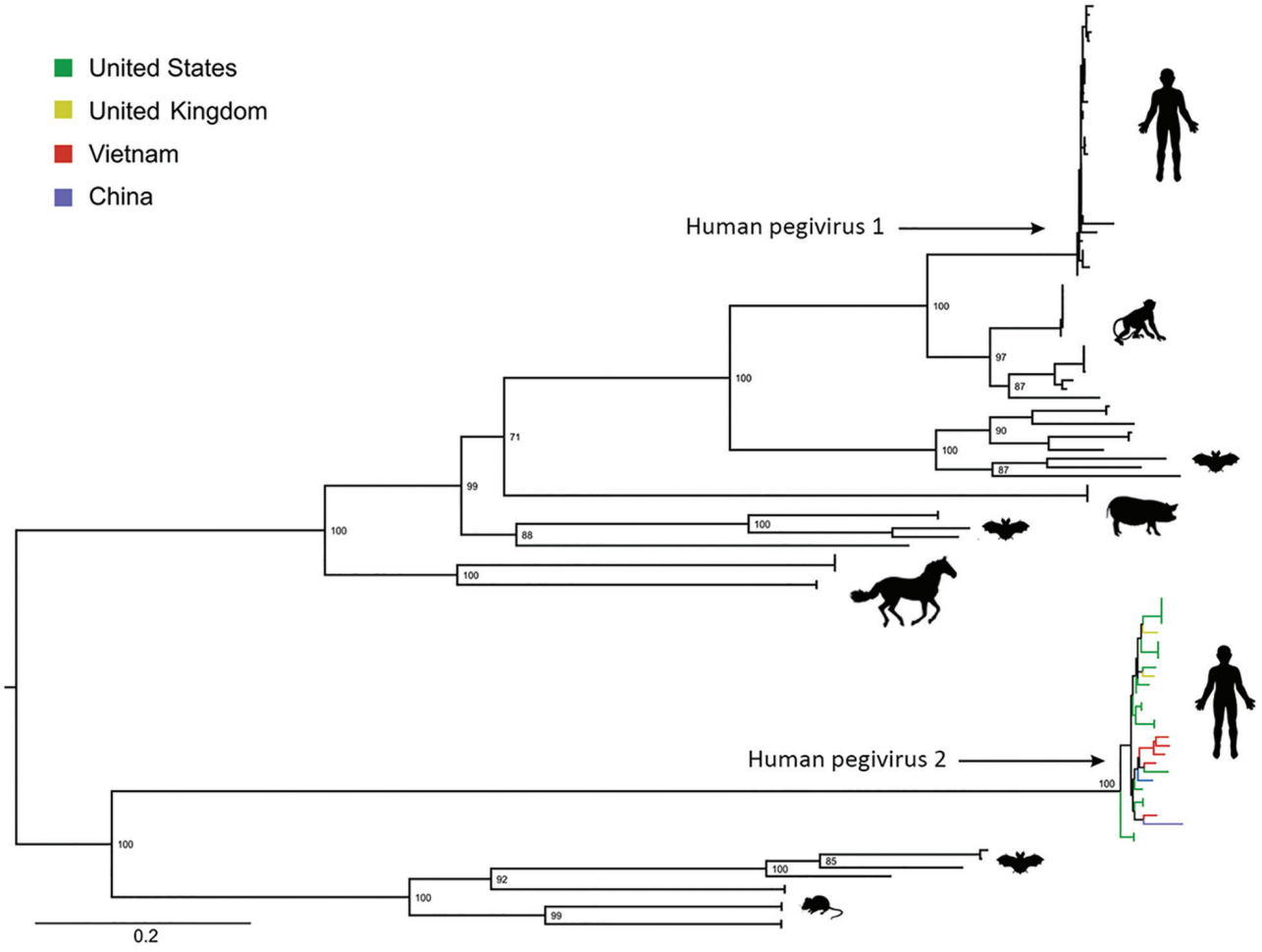


Figure. Maximum-likelihood phylogenetic tree of amino acid sequences of coding sequences of human pegivirus 2 strains from Vietnam compared with global strains and other pegiviruses. We used the general matrix with empirical amino acid frequencies, a gamma distribution of 4 rates, and invariant sites, as suggested by IQ TREE (http://www.iqtree.org), to reconstruct the phylogenetic trees. We assessed support for individual nodes using a bootstrap procedure of 10,000 replicates. Scale bar indicates amino acid substitutions per site.

Medicine Research Unit, Bangkok); Achara Laongnualpanich (Chiang Rai Prachanukroh Hospital, Chiang Rai, Thailand); Prapit Teparrakkul and Pramot Srisamang (Sappasithiprasong Hospital, Ubon Ratchathani, Thailand); Claire Ling (Shoklo Malaria Research Unit, Mae Sot, Thailand); Phan Huu Phuc and Le Thanh Hai (National Hospital of Pediatrics, Hanoi, Vietnam); Nguyen Van Kinh (National Hospital of Tropical Diseases, Hanoi); Bui Duc Phu (Hue Central Hospital, Hue, Vietnam); Nguyen Thanh Hung and Tang Chi Thuong (Children's Hospital 1, Ho Chi Minh City, Vietnam); Ha Manh Tuan (Children's Hospital 2, Ho Chi Minh City); Lam Minh Yen and Nguyen Van Vinh Chau (Hospital for Tropical Diseases, Ho Chi Minh City); Guy Thwaites, Heiman Wertheim, Le Van Tan, Motiur Rahman, and H. Rogier van Doorn (Oxford University Clinical Research Unit, Ho Chi Minh City); and Chuen-Yen Lau (National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA).

Acknowledgments

We thank Le Kim Thanh for her logistic support. We thank the patients for their participation in this study.

This study was funded by National Institute of Allergy and Infectious Diseases, US National Institutes of Health, and the Wellcome Trust of Great Britain (106680/B/14/Z and 204904/Z/16/Z). X.D. and E.D. were supported by the Blood Systems Research Institute and National Heart, Lung, and Blood Institute (grant no. R01 HL105770).

About the Author

Ms. Nguyen is a PhD student in life science at Open University, Milton Keynes, United Kingdom. Her research interests are virus discovery and evolution of emerging pathogens such as enteroviruses.

References

1. Berg MG, Lee D, Coller K, Frankel M, Aronsohn A, Cheng K, et al. Discovery of a novel human pegivirus in blood associated with hepatitis C virus co-infection. PLoS Pathog. 2015; 11:e1005325. http://dx.doi.org/10.1371/journal.ppat.1005325
2. Kapoor A, Kumar A, Simmonds P, Bhuva N, Singh Chauhan L, Lee B, et al. Virome analysis of transfusion recipients reveals a novel human virus that shares genomic features with hepaciviruses and pegiviruses. MBio. 2015;6:e01466–15. http://dx.doi.org/10.1128/mBio.01466-15
3. Wang H, Wan Z, Xu R, Guan Y, Zhu N, Li J, et al. A novel human pegivirus, HPgV-2 (HHpgV-1), is tightly associated with hepatitis C virus (HCV) infection and HCV/human immunodeficiency virus type 1 coinfection. Clin Infect Dis. 2018;66:29–35. http://dx.doi.org/10.1093/cid/cix748
4. Wang H, Wan Z, Sun Q, Zhu N, Li T, Ren X, et al. Second human pegivirus in hepatitis C virus-infected and hepatitis C virus/HIV-1 co-infected persons who injected drugs, China. Emerg Infect Dis. 2018;24:908–11.
5. Kandathil AJ, Breitwieser FP, Sachithanandham J, Robinson M, Mehta SH, Timp W, et al. Presence of human hepegivirus-1 in a cohort of people who inject drugs. Ann Intern Med. 2017;167:1–7. http://dx.doi.org/10.7326/M17-0085
6. Southeast Asia Infectious Disease Clinical Research Network. Causes and outcomes of sepsis in southeast Asia: a multinational multicentre cross-sectional study. Lancet Glob Health. 2017; 5:e157–67. http://dx.doi.org/10.1016/S2214-109X(17)30007-4
7. Nguyen AT, Tran TT, Hoang VMT, Nghiem NM, Le NNT, Le TTM, et al. Development and evaluation of a non-ribosomal random PCR and next-generation sequencing based assay for detection and sequencing of hand, foot, and mouth disease pathogens. Virol J. 2016;13:125. http://dx.doi.org/10.1186/s12985-016-0580-9
8. Frankel M, Forberg K, Coller KE, Berg MG, Hackett J Jr, Cloherty G, et al. Development of a high-throughput multiplexed real time RT-PCR assay for detection of human pegivirus 1 and 2. J Virol Methods. 2017;241:34–40. http://dx.doi.org/10.1016/j.jviromet.2016.12.013
9. Ladner JT, Beitzel B, Chain PS, Davenport MG, Donaldson EF, Frieman M, et al.; Threat Characterization Consortium. Standards for sequencing viral genomes in the era of high-throughput sequencing.. VIROME: a standard operating procedure for analysis of viral metagenome sequences. MBio. 2014;5:e01360–14.
10. Bijvand Y, Aghasadeghi MR, Sakhaee F, Pakzad P, Vaziri F, Saraji AA, et al. First detection of human hepegivirus-1 (HHpgV-1) in Iranian patients with hemophilia. Sci Rep. 2018;8:5036. http://dx.doi.org/10.1038/s41598-018-23490-4
11. Barclay ST, Cooke GS, Holtham E, Gauthier A, Schwarzbard J, Atanasov P, et al. A new paradigm evaluating cost per cure of HCV infection in the UK. Hepatol Med Policy. 2016;1:2. https://doi.org/10.1186/s41124-016-0002-z
12. Ireland G, Delpech V, Kirwan P, Croxford S, Lattimore S, Sabin C, et al. Prevalence of diagnosed HIV infection among persons with hepatitis C virus infection: England, 2008–2014. HIV Med. 2018 Jul 26 [Epub ahead of print].

Address for correspondence: To Anh Nguyen or Le Van Tan, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam; email: anhnt@oucru.org or tanlv@oucru.org

Viral Metagenomic Analysis of Cerebrospinal Fluid from Patients with Acute Central Nervous System Infections of Unknown Origin, Vietnam

Nguyen To Anh, Le Nguyen Truc Nhu, Nguyen Thi Thu Hong, Tran My Phuc, Pham Thi Thanh Tam, Dang Thao Huong, Tran Tuan Anh, Xutao Deng, Ho Dang Trung Nghia, Tran Thua Nguyen, Nguyen Van Hung, Nguyen Dac Thuan, Pham Thi Hong Phuong, Nguyen Van Vinh Chau, Stephen Baker, Eric Delwart, Guy Thwaites, Le Van Tan, for the VIZIONS Consortium[1]

Central nervous system (CNS) infection is a serious neurologic condition, although the etiology remains unknown in >50% of patients. We used metagenomic next-generation sequencing to detect viruses in 204 cerebrospinal fluid (CSF) samples from patients with acute CNS infection who were enrolled from Vietnam hospitals during 2012–2016. We detected 8 viral species in 107/204 (52.4%) of CSF samples. After virus-specific PCR confirmation, the detection rate was lowered to 30/204 (14.7%). Enteroviruses were the most common viruses detected (n = 23), followed by hepatitis B virus (3), HIV (2), molluscum contagiosum virus (1), and gemycircularvirus (1). Analysis of enterovirus sequences revealed the predominance of echovirus 30 (9). Phylogenetically, the echovirus 30 strains belonged to genogroup V and VIIb. Our results expanded knowledge about the clinical burden of enterovirus in Vietnam and underscore the challenges of identifying a plausible viral pathogen in CSF of patients with CNS infections.

Worldwide, the annual incidence of acute encephalitis in nonoutbreak settings during 1983–2000 ranged from 0.07 to 12.6 cases/100,000 population (*1*). According to the World Health Organization, meningitis caused 379,000 deaths and encephalitis caused 150,000 deaths globally in 2015 (*2*). As a consequence, central nervous system (CNS) infection is a leading cause of years lived with disability in low-income countries (*3*).

More than 100 known pathogens can cause CNS infections (*1*). However, the distribution of CNS infection pathogens is geographically dependent and has been shaped by the emergence of novel viruses. In Asia, Nipah virus and enterovirus A71 have been recognized as emerging neurotropic pathogens over the past few decades. In 1999, West Nile virus arrived in the United States and since then has established endemic circulation (*4*).

Despite recent advances in molecular diagnostics, especially sensitive virus-specific PCR, encephalitis cases of unknown origin remain a substantial problem. Worldwide, ≈50% of patients with CNS infections have no etiology identified (*1*,*5*,*6*).

Over the past decade, metagenomic next-generation sequencing (mNGS) has emerged as a sensitive hypothesis-free approach for detection of pathogens (especially viruses) in clinical samples (*7*). However, in resource-limited settings like Southeast Asia and Vietnam, a limited number of mNGS studies

Author affiliations: Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam (N.T. Anh, L.N.T. Nhu, N.T.T. Hong, T.M. Phuc, P.T.T. Tam, D.T. Huong, T.T. Anh, H.D.T. Nghia, G. Thwaites, L.V. Tan); Hue Central Hospital, Hue City, Vietnam (T.T. Nguyen); Dak Lak General Hospital, Ban Me Thuot City, Vietnam (N.V. Hung); Khanh Hoa General Hospital, Nha Trang City, Vietnam (N.D. Thuan); Dong Thap General Hospital, Dong Thap, Vietnam (P.T.H. Phuong); Vitalant Research Institute, San Francisco, California, USA (X. Deng, E. Delwart); University of California Department of Laboratory Medicine, San Francisco (X. Deng, E. Delwart); Pham Ngoc Thach University, Ho Chi Minh City (H.D.T. Nghia); Hospital for Tropical Diseases, Ho Chi Minh City (N.V.V. Chau); University of Cambridge Institute of Therapeutic Immunology and Infectious Disease, Cambridge, UK (S. Baker); Centre for Tropical Medicine, Nuffield Department of Medicine, University of Oxford, Oxford, UK (S. Baker, G. Thwaites)

DOI: https://doi.org/10.3201/eid2701.202723

[1]Members are listed at the end of this article.

examining known and unknown viruses in cerebrospinal fluid (CSF) samples from patients with CNS infections have been conducted, even though in this tropical region of the world, novel viruses are likely to emerge (P. Zhou et al., unpub. data, https://doi.org/10.1101/2020.01.22.914952), and diverse CNS infection pathogens have been documented. Collectively, improving our knowledge about viral causes of CNS infections is essential for clinical management and development of intervention strategies. In this study, by using a mNGS approach, we set out to search for known and unknown viruses in CSF samples collected from patients in Vietnam with CNS infections of unknown causes who were enrolled in a hospital-based surveillance study conducted during 2012–2016.

Materials and Methods

Clinical Study and Selection of CSF Samples for mNGS Analysis

The study used CSF samples collected from patients with suspected CNS infection enrolled in a hospital-based surveillance program conducted in Vietnam during December 2012–October 2016 (*5*). The study was conducted as part of the Vietnam Initiative on Zoonotic Infections (VIZIONS) project (*5*), and patient recruitment was carried out at 7 provincial hospitals across Vietnam. After collection, as per the study protocol, all CSF samples were tested for a range of pathogens by using the diagnostic work-up of the clinical study (Appendix Table 1, https://wwwnc.cdc.gov/EID/article/27/1/20-2723-App1.pdf). The remaining volume of the CSF samples were stored at −80°C for further testing.

We focused our metagenomic analysis on patients of unknown origin from 4 provincial hospitals in central (Hue and KhanhHoa), highland (DakLak), and southern (DongThap) Vietnam (Figure 1), representing 3 distinct geographic areas in Vietnam. To increase the chance of detecting a virus in the CSF samples, we only selected patients with CSF leukocyte counts ≥5 cells/mm^3 and an illness duration ≤5 days.

mNGS Assay

mNGS assay was carried out as previously described (*8*). Before viral nucleic acid (NA) isolation, 100 µL of each CSF sample was treated with Turbo DNase (Ambion, Life Technology, ThermoFisher, https://www.thermofisher.com) and RNase I enzyme (Ambion). Then viral NA was isolated using a QIAamp viral RNA kit (QIAGEN GmbH, https://www.qiagen.com), and recovered in 50 µL of elution buffer provided with the extraction kit. Double-stranded DNA was synthesized from the isolated viral NA by using a set of 96 nonribosomal primers (FR26RV–Endoh primers) and then was randomly amplified by using the FR20RV primer (5′-GCCGGAGCTCTGCAGATATC-3′). Finally, the amplified product was subjected to a library preparation step by using Nextera XT sample preparation kit (Illumina, https://www.illumina.com), following the manufacturer's instructions, and sequenced by using a MiSeq reagent kit, version 3 (600 cycles) (Illumina) in a MiSeq platform (Illumina).

mNGS Data Analysis

Potential viral reads were identified by using an in-house viral metagenomic pipeline running on a 36-node Linux cluster as described previously (*9*). In

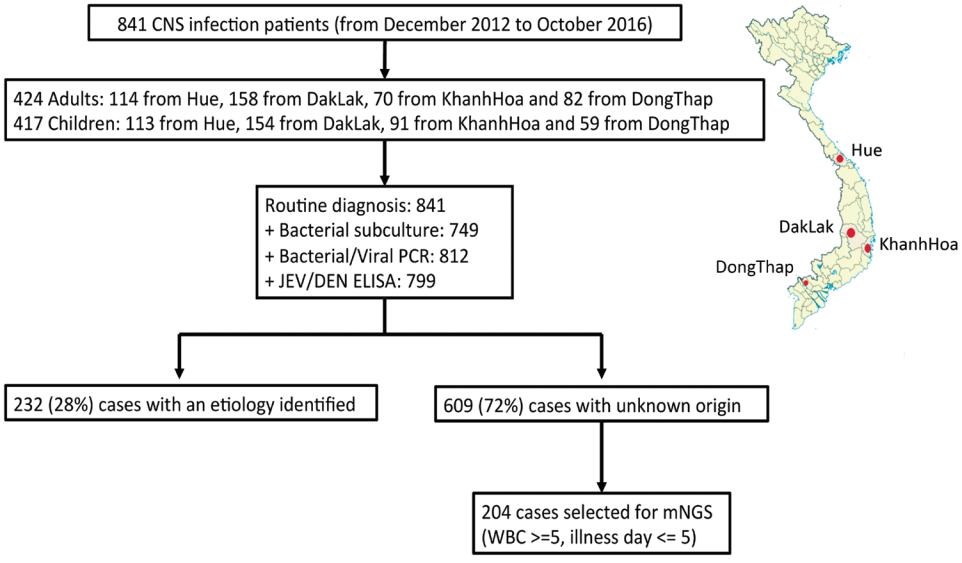


Figure 1. Flowchart overview of diagnostic results for study of patients with suspected central nervous system infections admitted to 4 of 7 provincial hospitals, Vietnam, December 2012–October 2016. Inset map indicates places where samples were collected (red dots).

brief, after duplicate reads and reads belonging to human or bacterial genomes were filtered out, the remaining reads were assembled de novo. The resulting contigs and singlet reads were then aligned against a customized viral proteome database by using an approach based on BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Next, the candidate viral reads were aligned against a nonredundant nonvirus protein database to remove any false-positive reads (i.e., reads with expected values higher than those against viral protein databases). Any virus-like sequence with an expected value ≤0.00001 was considered a significant hit. Finally, a reference-based mapping approach (Genious 8.1.5; Biomatters, https://www.geneious.com) was used to assess the levels of identity and genome coverage of the corresponding viruses.

PCR Confirmatory Testing of mNGS Results

PCR assays were conducted to confirm mNGS hits for each specific virus identified from the viral metagenomic pipeline. Depending on availability of CSF, the PCR confirmations were performed either on leftover NA or newly extracted NA. A viral mNGS result was considered positive only if it was subsequently confirmed by PCR analysis of the original NA samples. The nucleotide sequences of primers and probes used for PCR confirmatory testing are shown in Appendix Table 2 (*8*).

Serotype Identification and Phylogenetic Analysis

For enterovirus serotype determination based on the obtained sequences generated by viral mNGS, we used a publicly available genotyping tool (*10*). To determine the relationship between enterovirus strains we sequenced and global strains, we first performed pairwise alignment by using the ClustalW tool in Geneious 8.1.5, and then reconstructed a maximum-likelihood phylogenetic tree by using IQ Tree 1.4.3 (*11*). A similar phylogenetic approach was used for other viruses. The generated sequences of this study were submitted to GenBank (accession no. PRJNA561465).

Ethics

The study was approved by the corresponding institutional review broad of local hospitals in Vietnam, where the patients were enrolled, and the Oxford Tropical Ethics Committee. Informed consent was obtained from each study participant or a legal guardian.

Results

CSF Samples Available for mNGS Analysis

From the clinical study described previously, a total of 841 patients were enrolled from Hue, Khanh Hoa, Dak Lak, or Dong Thap provincial hospitals. Of these, 609/841 (72%) patients had no etiology identified. The etiologic profiles of the patients in

Table. Baseline characteristics and clinical data of patients with acute central nervous system infections enrolled for mNGS analysis of CSF samples, Vietnam, December 2012–October 2016*

Characteristic	Patients with unknown cause enrolled for mNGS, n = 204	Patients with mNGS negative, n = 174	Patients with enterovirus detected, n = 23	p value†
Sex				
M	135 (66)	114 (65.5)	15 (65)	
F	69 (34)	60 (34.5)	8 (35)	
Age, y, median (range)	20.5 (0–92)	24 (0–92)	13 (2–27)	0.005
Location				
Hue	37 (18)	28 (16)	9 (39)	
Dak lak	98 (48)	87 (50)	10 (43.5)	
Khanh Hoa	28 (14)	22 (13)	4 (17.5)	
Dong Thap	41 (20)	37 (21)	0	
3-d fever (at enrollment or preceding 3 d)				
Fever	148 (72.5)	126 (72.4)	17 (74)	0.054
Temperature, C°, median (range)	39 (37.5–42.0)	39 (37.5–42.0)	38.5 (38.0–40.5)	
Fever with unknown temperature	29 (14.2)	22 (12.6)	6 (26)	
No fever	20 (9.8)	19 (11)	0	
Unknown	7 (3.5)	7 (4)	0	
Outcome				
Death or discharge to die	22 (11)	22 (12.6)	0	
Discharge with complete recovery	108 (53)	86 (49.4)	18 (78.3)	
Discharge with incomplete recovery	35 (17)	31 (17.8)	2 (8.7)	
Transfer to another hospital	34 (16.5)	30 (17.2)	3 (13)	
Other (patient request)	3 (1.5)	3 (1.7)	0	
Unknown	2 (1)	2 (1.3)	0	
CSF white cells, cells/mm^3 median (min–max)	88.5 (5–40,000)	71.5 (5–40,000)	110 (8–1200)	0.343

*Values are no. (%) except as indicated, CSF, cerebrospinal fluid; mNGS, metagenomic next-generation sequencing.
†Statistical comparisons were performed for groups of patients with mNGS-negative results and enterovirus detected, by Mann-Whiney test.

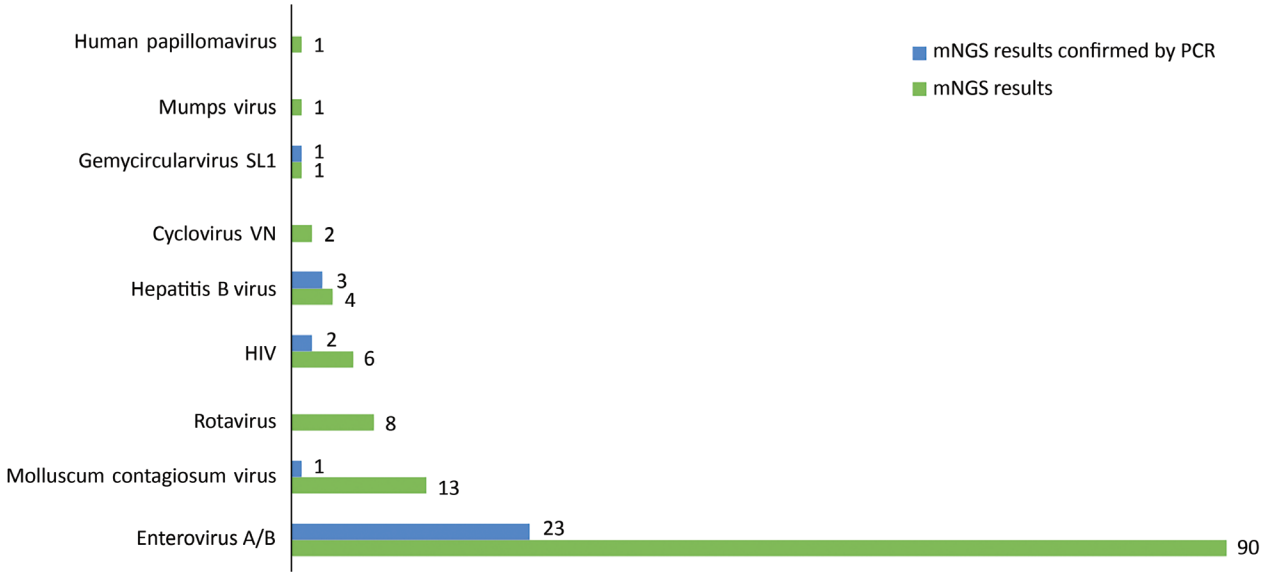


Figure 2. Number of cerebrospinal fluid samples with detected viruses by metagenomic next-generation sequencing and then confirmed by virus-specific PCR or reverse-transcription PCR, Vietnam, December 2012–October 2016. Samples were collected from patients with suspected central nervous system infection. For human papillomavirus, confirmatory testing was not performed because of the unavailability of a PCR assay.

whom a pathogen was detected will be reported separately. Of the patients in whom a pathogen was not identified, 204 met our selection criteria, and their CSF samples were included for viral mNGS analysis (Figure 1).

Baseline Characteristics of the Included Patients

The baseline characteristics and outcome of the 204 study patients are described in Table 1. Male patients were predominant. A substantial proportion of the patients were seriously ill; fatal outcome was recorded in 22 (11%), whereas incomplete recovery was recorded in 17% (n = 35) and deterioration (reflected by being transferred to other hospitals) in 16.5% (n = 34).

General Description of mNGS Results

A total of 204 CSF samples were subjected to 3 NGS runs, and 108 million reads were obtained (median number of reads per sample 445,412 [range 430–908,890]). Of these, viral reads accounted for 0.64% (n = 692,731; median number of reads per sample 2,001 [range 4–268,933]). Excluding common contaminants and commensal viruses such as torque teno virus, which are not reported in this article, sequences related to a total of 8 distinct viral species were identified in 107/204 (52.4%) patients. These viruses are either known to be infectious to humans (e.g., enteroviruses, rotavirus, molluscum contagiosum virus [MCV], human papillomavirus, HIV, and hepatitis B virus [HBV]) or are without evidence of human infections besides previous detection in sterile human samples (e.g., cyclovirus-VN and gemycircularvirus) (Figure 2).

mNGS Result Assessment by Specific PCR Analysis

After virus-specific PCR confirmatory testing, the proportion of patients in whom a virus was found by mNGS was reduced from 53% (108/204) to 14.7% (30/204). Accordingly, the number of virus species was reduced from 8 to 5 (Figure 2); enteroviruses were the most common virus detected, accounting for 11.3% (23/204) of the included patients, followed by HBV (n = 3), HIV (n = 2), gemycircularvirus, and MCV (1 each) (Figure 2). Because of the focus of our study and the unavailability of the PCR assays, confirmatory testing for human papillomavirus was not performed.

Characteristics of the 23 Enterovirus-Infected Patients

All 23 enterovirus-infected patients were admitted to hospitals from the central or highland areas (Table), and none were from Dong Thap Province. Male patients were slightly predominant, accounting for

56%. Notably, the enterovirus-infected patients were younger than those who were mNGS-negative (Table). At discharge, incomplete recovery or transfer to other hospitals because of disease deterioration were recorded in 21.7% (Table).

Enterovirus cases were not detected during January 2015–December 2016. During 2013 and 2014, two main peaks were observed during March–July and September–December (Figure 3, panel A); cases from Dak Lak and Khanh Hoa contributed to the first peak (Figure 3, panels B and C), and cases from Khanh Hoa and Hue contributed to the second (Figure 3, panels C and D). The general baseline characteristics of patients with HBV, HIV, gemycircularvirus, and MCV are shown in Appendix Table 3.

Genetic Characterization of Enteroviruses and Gemycircularvirus

mNGS generated sufficient sequence information for an enterovirus serotyping assessment in 11/23 cases. Subsequently, results of serotyping analysis based on the NGS sequences showed that echovirus 30 (E30)

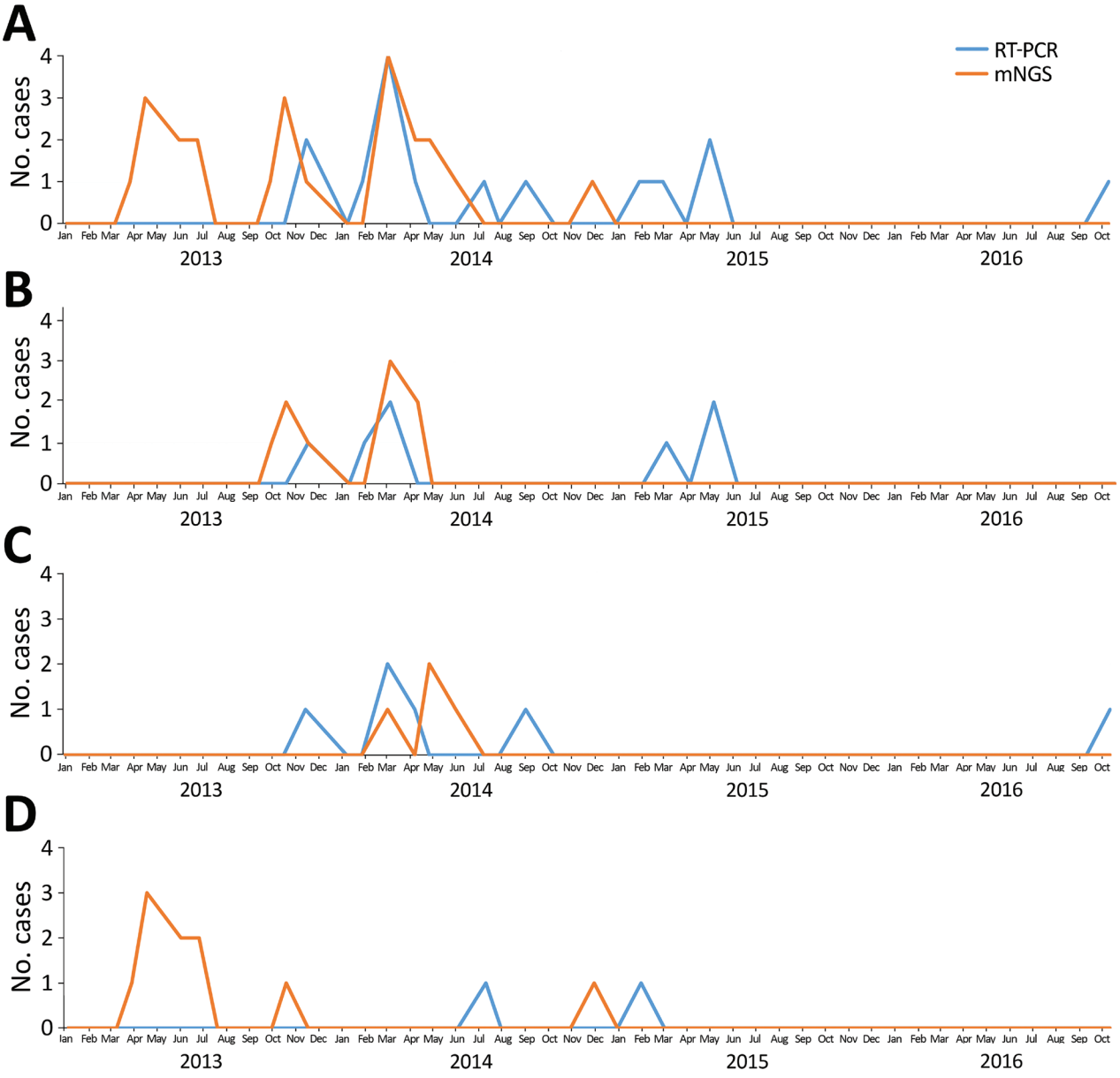


Figure 3. Temporal distribution of enterovirus cases detected from cerebrospinal fluid samples of patients with suspected central nervous system infection by metagenomic next-generation sequencing and RT-PCR, Vietnam, December 2012–October 2016. Enterovirus RT-PCR results were obtained from the original study. RT-PCR, reverse transcription PCR. A) Combined data from 3 provinces; B) data from Hue province; C) data from Khanh Hoa province; D) data from Dak Lak province.

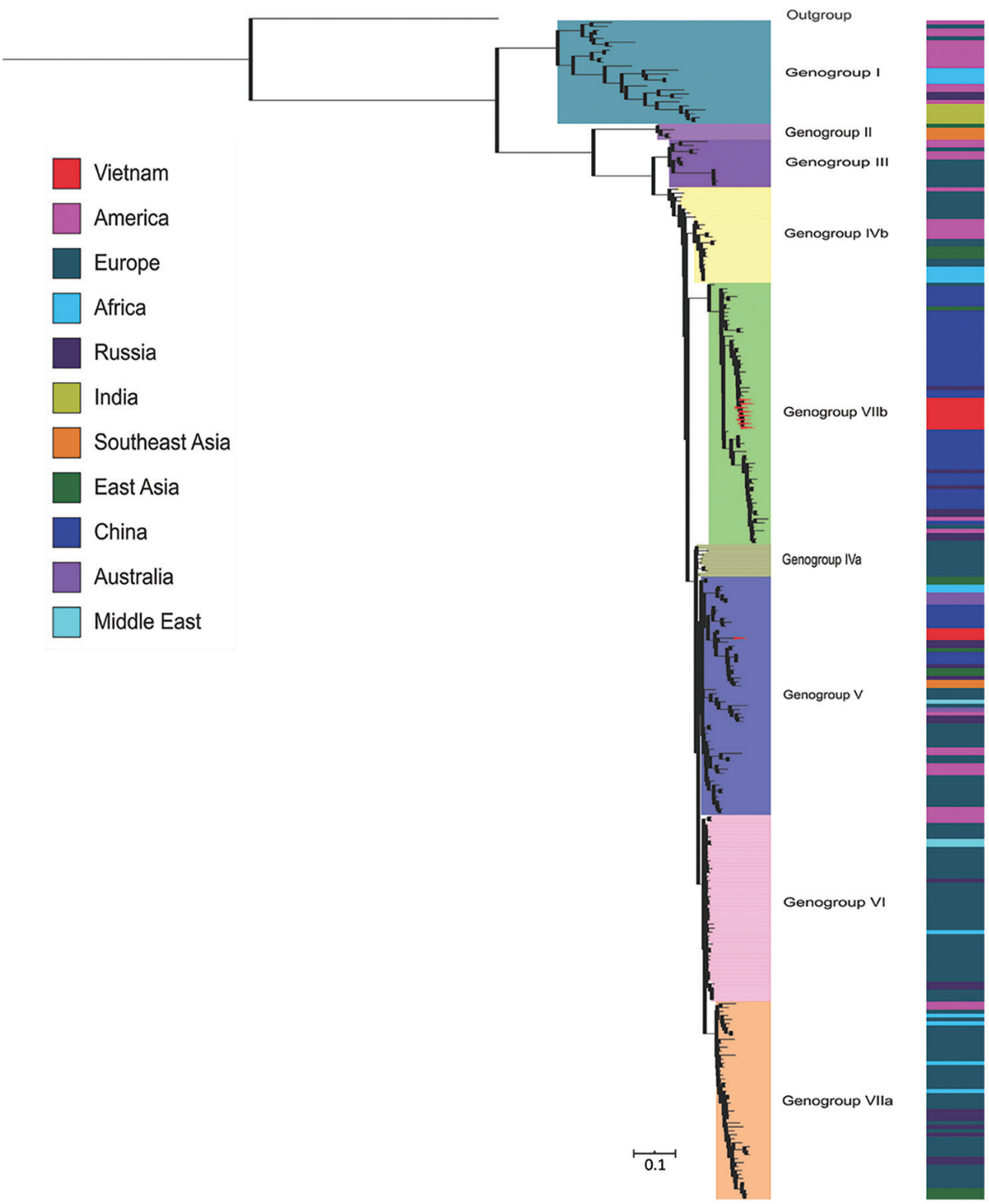


Figure 4. Phylogenetic tree of 298 complete viral protein 1 sequences of echovirus 30 (876 nt) isolated from cerebrospinal fluid samples of patients with suspected central nervous system infection, Vietnam, December 2012–October 2016. The inner color strip indicates 7 genogroups. The outer color strip indicates different countries of echovirus 30 isolates included in the tree. The outgroup is echovirus 21 Farina. The E30 sequences generated by metagenomic next-generation sequencing ae highlighted in red.

was the most common serotype detected (n = 9, 39% of enteroviruses), followed by enterovirus A71 and enterovirus B80 (1 each, 4.3%). Phylogenetically, the 9 E30 strains sequenced in our study belonged to 2 distinct genogroups, V and VIIb, and showed close relationship with E30 strains circulating in Russia and elsewhere in Asia, including China (Figure 4).

In additional to enterovirus sequences, a gemycircularvirus genome was obtained from a 12-year-old boy. Phylogenetic analysis revealed that this gemycircularvirus strain was closely related to a gemycircularvirus species previously found in CSF sample from a patient with a CNS infection of unknown origin in Sri Lanka (*12*); the level of amino acid identities between the 2 strains were 98.79% for replication-coding sequences and 99.3% for capsid protein–coding sequences.

Discussion

We describe a viral mNGS investigation characterizing the human virome in CSF of 204 patients in Vietnam with suspected CNS infection of unknown origin. We successfully detected 4 human viral pathogens (enteroviruses, HIV, HBV, and MCV) and 1 virus species (gemycircularvirus) of unknown tropism and pathogenicity in a total of 30 (14.7%) patients. Most patients therefore remained without a known etiology, underscoring the ongoing challenge in identifying a plausible viral pathogen in CSF of patients with CNS infections.

Enteroviruses were the most common viruses, found in 11.3% (23/204) of all analyzed patients (Figure 2), most of whom were children and young adults. This age distribution of enterovirus-infected patients is consistent with observational data from a previous report from Vietnam (*6*), although the median age was slightly higher compared with data from other countries (*13*,*14*). Geographically, all the enterovirus-infected patients were admitted to hospitals from central and highland Vietnam, and none was from southern Vietnam. The underlying mechanism determining this observed spatial pattern of enterovirus-positive cases in this study remains unknown. Our sampling timescale perhaps was not long enough to capture the circulation of enteroviruses in Dong Thap Province. Enteroviruses were previously reported as a leading cause of CNS infection across central and southern Vietnam (*6*,*15*,*16*). Collectively, our findings suggest that reverse transcription PCR (RT-PCR) testing for enteroviruses should be considered in children and young adults with CNS infections.

Of the detected enteroviruses, E30 was the most common serotype. E30 is a well-known pathogen of pediatric aseptic meningitis worldwide (*17*). Phylogenetically, at global scale, E30 belongs to 2 different lineages with distinct patterns of circulation and spread, 1 with a global distribution and the other with geographic restriction within Asia (*17*). The cocirculation of 2 E30 lineages in Vietnam suggests that E30 was imported into Vietnam on at least 2 occasions. Our analyses thus also contribute to the body of knowledge about the genetic diversity of E30 strains circulating in Vietnam.

The detection of bloodborne viruses such as HBV and HIV is unlikely to have a direct link with patients' neurologic symptoms, although HBV has previously been reported in CSF of patients with CNS infections of unknown origin (*18*). The detection of HIV in CSF might have been a consequence of traumatic tap occurring during the lumbar puncture, as reflected by the high number of red blood cells in 1 of 2 HIV-positive CSF samples (data not shown). However, neuroinvasion of HIV has also been reported (*19*). Likewise, the pathogenic potential of a gemycircularvirus genome requires further investigation, although the detection of the gemycircularvirus genome in CSF has been reported in several papers (*12*,*18*,*20*). The detection of MCV and papillomavirus in CSF might result from contamination of viral skin flora during lumbar puncture.

Similar to previous reports about discrepancy between mNGS and conventional diagnostic testing (*8*,*18*,*21*), our observations found that most mNGS-positive results were not confirmed by subsequent viral RT-PCR assays, especially the sensitive enterovirus-specific RT-PCR with a limit of detection of ≈9 copies/reaction (*22*). Such results could be attributable to bleedover (also called index hopping) of indices from reads of 1 sample into reads of another sample co-sequenced on the same Illumina run (R. Sinha et al., unpub. data, https://doi.org/10.1101/125724). Applying double indexes, which was not used in our study, has been shown to substantially reduce, but not eliminate, the cross-contamination phenomenon between samples in the same run.

Our study has some limitations. First, as outlined previously, we did not employ a double unique index combination strategy per sample as part of the sequencing procedure. The well-known index hopping phenomenon possibly explains the high discrepancy between confirmatory PCR and mNGS results (*21*,*23*,*24*) and emphasizes the usefulness of dual indexing and including no template controls. As such, we pragmatically chose to verify our mNGS by performing specific PCR on original materials. Second, the DNase treatment step in our assay meant to reduce cellular DNA concentration in CSF

might reduce the sensitivity of mNGS for the detection of DNA viruses such as herpes simplex virus (*25*,*26*). Third, some of the non–PCR-confirmed viral sequences likely originated from contamination of reagents, which is a lingering problem for mNGS (*27*,*28*).

In summary, our results emphasize that mNGS could detect a broad range of viral nucleic acids in CSF. In spite of extensive investigation, establishing the etiology in many patients with CNS infections remains a challenge. However, our findings indicate that enteroviruses are important causes of viral CNS infections in Vietnam and thus should be considered in the differential diagnosis among young patients with CNS infections.

VIZIONS Consortium members: from the Oxford University Clinical Research Unit, Bach Tuan Kiet, Stephen Baker, Alessandra Berto, Maciej F. Boni, Juliet E. Bryant, Bui Duc Phu, James I. Campbell, Juan Carrique-Mas, Dang Manh Hung, Dang Thao Huong, Dang Tram Oanh, Jeremy N. Day, Dinh Van Tan, H. Rogier van Doorn, Duong An Han, Jeremy J. Farrar, Hau Thi Thu Trang, Ho Dang Trung Nghia, Hoang Bao Long, Hoang Van Duong, Huynh Thi Kim Thu, Lam Chi Cuong, Le Manh Hung, Le Thanh Phuong, Le Thi Phuc, Le Thi Phuong, Le Xuan Luat, Luu Thi Thu Ha, Ly Van Chuong, Mai Thi Phuoc Loan, Behzad Nadjm, Ngo Thanh Bao, Ngo Thi Hoa, Ngo Tri Tue, Nguyen Canh Tu, Nguyen Dac Thuan, Nguyen Dong, Nguyen Khac Chuyen, Nguyen Ngoc An, Nguyen Ngoc Vinh, Nguyen Quoc Hung, Nguyen Thanh Dung, Nguyen Thanh Minh, Nguyen Thi Binh, Nguyen Thi Hong Tham, Nguyen Thi Hong Tien, Nguyen Thi Kim Chuc, Nguyen Thi Le Ngoc, Nguyen Thi Lien Ha, Nguyen Thi Nam Lien, Nguyen Thi Ngoc Diep, Nguyen Thi Nhung, Nguyen Thi Song Chau, Nguyen Thi Yen Chi, Nguyen Thieu Trinh, Nguyen Thu Van, Nguyen Van Cuong, Nguyen Van Hung, Nguyen Van Kinh, Nguyen Van Minh Hoang, Nguyen Van My, Nguyen Van Thang, Nguyen Van Thanh, Nguyen Van Vinh Chau, Nguyen Van Xang, Pham Ha My, Pham Hong Anh, Pham Thi Minh Khoa, Pham Thi Thanh Tam, Pham Van Lao, Pham Van Minh, Phan Van Be Bay, Maia A. Rabaa, Motiur Rahman, Corinne Thompson, Guy Thwaites, Ta Thi Dieu Ngan, Tran Do Hoang Nhu, Tran Hoang Minh Chau, Tran Khanh Toan, Tran My Phuc, Tran Thi Kim Hong, Tran Thi Ngoc Dung, Tran Thi Thanh Thanh, Tran Thi Thuy Minh, Tran Thua Nguyen, Tran Tinh Hien, Trinh Quang Tri, Vo Be Hien, Vo Nhut Tai, Vo Quoc Cuong, Voong Vinh Phat, Vu Thi Lan Huong, Vu Thi Ty Hang, and Heiman Wertheim; from the Centre for Immunity, Infection, and Evolution, University Of Edinburgh, Edinburgh, Scotland, UK, Carlijn Bogaardt, Margo Chase-Topping, Al Ivens, Lu Lu, Dung Nyugen, Andrew Rambaut, Peter Simmonds, and Mark Woolhouse; from The Wellcome Trust Sanger Institute, Hinxton, UK, Matthew Cotten, Bas B. Oude Munnink, Paul Kellam, and My Vu Tra Phan; from the Laboratory of Experimental Virology, Department of Medical Microbiology, Center for Infection and Immunity Amsterdam (CINIMA), Academic Medical Center of the University of Amsterdam, Amsterdam, the Netherlands, Martin Deijs, Lia van der Hoek, Maarten F. Jebbink, and Seyed Mohammad Jazaeri Farsani; and from Metabiota, California, USA, Karen Saylors and Nathan Wolfe.

Acknowledgments

We would like to thank Le Thi Kim Thanh for her logistic support and the patients for their participation in this study.

The study was funded by the Wellcome Trust of Great Britain (awards nos. WT/093724 and 106680/B/14/Z [to the Oxford University Clinical Research Unit in Vietnam], 100087/Z/12/Z [to S.B.] and 204904/Z/16/Z [to L.V.T.]), and the Royal Society (grant no. 098511/Z/12/Z [to S.B.]). X.D. and E.D. were supported by the Blood Systems Research Institute and the National Heart, Lung, and Blood Institute (grant no. R01 HL105770).

About the Author

Ms. Nguyen is a PhD student in life science at Open University, Milton Keynes, UK. Her research interests are virus discovery and evolution of emerging pathogens such as enteroviruses.

References

1. Granerod J, Tam CC, Crowcroft NS, Davies NWS, Borchert M, Thomas SL. Challenge of the unknown. A systematic review of acute encephalitis in non-outbreak situations. Neurology. 2010;75:924–32. https://doi.org/10.1212/WNL.0b013e3181f11d65
2. Feigin VL, Krishnamurthi RV, Theadom AM, Abajobir AA, Mishra SR, Ahmed MB, et al.; GBD 2015 Neurological Disorders Collaborator Group. Global, regional, and national burden of neurological disorders during 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Neurol. 2017;16:877–97. https://doi.org/10.1016/S1474-4422(17)30299-5
3. World Health Organization. Neurological disorders: public health challenges. 2006 [cited 2017 Nov 11]. https://www.who.int/mental_health/publications/neurological_disorders_ph_challenges
4. Sejvar JJ. West nile virus: an historical overview. Ochsner J. 2003;5:6–10.
5. Rabaa MA, Tue NT, Phuc TM, Carrique-Mas J, Saylors K, Cotten M, et al. The Vietnam Initiative on Zoonotic Infections (VIZIONS): a strategic approach to studying emerging zoonotic infectious diseases. EcoHealth. 2015;12:726–35. https://doi.org/10.1007/s10393-015-1061-0

6. Ho Dang Trung N, Le Thi Phuong T, Wolbers M, Nguyen Van Minh H, Nguyen Thanh V, Van MP, et al.; VIZIONS CNS Infection Network. Aetiologies of central nervous system infection in Viet Nam: a prospective provincial hospital-based descriptive surveillance study. PLoS One. 2012;7:e37825. https://doi.org/10.1371/journal.pone.0037825
7. Brown JR, Bharucha T, Breuer J. Encephalitis diagnosis using metagenomics: application of next generation sequencing for undiagnosed cases. J Infect. 2018;76:225–40. https://doi.org/10.1016/j.jinf.2017.12.014
8. Anh NT, Hong NTT, Nhu LNT, Thanh TT, Lau C-Y, Limmathurotsakul D, et al. Viruses in Vietnamese patients presenting with community-acquired sepsis of unknown cause. J Clin Microbiol. 2019;57:e00386–19. https://doi.org/10.1128/JCM.00386-19
9. Aiemjoy K, Altan E, Aragie S, Fry DM, Phan TG, Deng X, et al. Viral species richness and composition in young children with loose or watery stool in Ethiopia. BMC Infect Dis. 2019;19:53. https://doi.org/10.1186/s12879-019-3674-3
10. Kroneman A, Vennema H, Deforche K, v d Avoort H, Peñaranda S, Oberste MS, et al. An automated genotyping tool for enteroviruses and noroviruses. J Clin Virol. 2011;51:121–5. https://doi.org/10.1016/j.jcv.2011.03.006
11. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74. https://doi.org/10.1093/molbev/msu300
12. Phan TG, Mori D, Deng X, Rajidrajith S, Ranawaka U, Fan Ng TF, et al. Small viral genomes in unexplained cases of human encephalitis, diarrhea, and in untreated sewage. Virology. 2015;482:98–104. https://doi.org/10.1016/j.virol.2015.03.011
13. Sun Y, Miao Z, Yan J, Gong L, Chen Y, Chen Y, et al. Sero-molecular epidemiology of enterovirus-associated encephalitis in Zhejiang Province, China, from 2014 to 2017. Int J Infect Dis. 2019;79:58–64. https://doi.org/10.1016/j.ijid.2018.11.002
14. Richter J, Tryfonos C, Christodoulou C. Molecular epidemiology of enteroviruses in Cyprus 2008–2017. PLoS One. 2019;14:e0220938. https://doi.org/10.1371/journal.pone.0220938
15. Taylor WR, Nguyen K, Nguyen D, Nguyen H, Horby P, Nguyen HL, et al. The spectrum of central nervous system infections in an adult referral hospital in Hanoi, Vietnam. PLoS One. 2012;7:e42099. https://doi.org/10.1371/journal.pone.0042099
16. Tan V, Thai H, Phu NH, Nghia HDT, Chuong LV, Sinh DX, et al. Viral aetiology of central nervous system infections in adults admitted to a tertiary referral hospital in southern Vietnam over 12 years. PLoS Negl Trop Dis. 2014;8:e3127. https://doi.org/10.1371/journal.pntd.0003127
17. Lema C, Torres C, Van der Sanden S, Cisterna D, Freire MC, Gómez RM. Global phylodynamics of Echovirus 30 revealed differential behavior among viral lineages. Virology. 2019;531:79–92. https://doi.org/10.1016/j.virol.2019.02.012
18. Schibler M, Brito F, Zanella MC, Zdobnov EM, Laubscher F, L'Huillier AG, et al. Viral sequences detection by high-throughput sequencing in cerebrospinal fluid of individuals with and without central nervous system disease. Genes (Basel). 2019;10:1–12. https://doi.org/10.3390/genes10080625
19. Zayyad Z, Spudich S. Neuropathogenesis of HIV: from initial neuroinvasion to HIV-associated neurocognitive disorder (HAND). Curr HIV/AIDS Rep. 2015;12:16–24. https://doi.org/10.1007/s11904-014-0255-3
20. Zhou C, Zhang S, Gong Q, Hao A. A novel gemycircular-virus in an unexplained case of child encephalitis. Virol J. 2015;12:197. https://doi.org/10.1186/s12985-015-0431-0
21. Wilson MR, Sample HA, Zorn KC, Arevalo S, Yu G, Neuhaus J, et al. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. N Engl J Med. 2019;380:2327–40. https://doi.org/10.1056/NEJMoa1803396
22. Thanh TT, Anh NT, Tham NT, Van HM, Sabanathan S, Qui PT, et al. Validation and utilization of an internally controlled multiplex Real-time RT-PCR assay for simultaneous detection of enteroviruses and enterovirus A71 associated with hand foot and mouth disease. Virol J. 2015;12:85. https://doi.org/10.1186/s12985-015-0316-2
23. Wilson MR, Fedewa G, Stenglein MD, Olejnik J, Rennick LJ, Nambulli S, et al. Multiplexed metagenomic deep sequencing to analyze the composition of high-priority pathogen reagents. mSystems. 2016;1:1–9. https://doi.org/10.1128/mSystems.00058-16
24. Yang J, Yang F, Ren L, Xiong Z, Wu Z, Dong J, et al. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. J Clin Microbiol. 2011;49:3463–9. https://doi.org/10.1128/JCM.00273-11
25. Nguyen TTH, Nguyen TA, Nguyen Thi Hoang M, Ho DTN, Le NNT. Tran tan T, et al. Performance of metagenomic next-generation sequencing for the diagnosis of viral meningoencephalitis in a resource limited setting. Open Forum Infect Dis. 2020;7:ofaa046. https://doi.org/10.1093/ofid/ofaa046
26. Perlejewski K, Popiel M, Laskus T, Nakamura S, Motooka D, Stokowy T, et al. Next-generation sequencing (NGS) in the identification of encephalitis-causing viruses: Unexpected detection of human herpesvirus 1 while searching for RNA pathogens. J Virol Methods. 2015;226:1–6. https://doi.org/10.1016/j.jviromet.2015.09.010
27. Asplund M, Kjartansdóttir KR, Mollerup S, Vinner L, Fridholm H, Herrera JAR, et al. Contaminating viral sequences in high-throughput sequencing viromics: a linkage study of 700 sequencing libraries. Clin Microbiol Infect. 2019;25:1277–85. https://doi.org/10.1016/j.cmi.2019.04.028
28. Holmes EC. Reagent contamination in viromics: all that glitters is not gold. Clin Microbiol Infect. 2019;25:1167–8. https://doi.org/10.1016/j.cmi.2019.06.019

Address for correspondence: Nguyen To Anh or Le Van Tan, Oxford University Clinical Research Unit, 764 VoVan Kiet, District 5, Ho Chi Minh City, Vietnam; email: anhnt@oucru.org or email: tanlv@oucru.org

Open Forum Infectious Diseases

MAJOR ARTICLE

Performance of Metagenomic Next-Generation Sequencing for the Diagnosis of Viral Meningoencephalitis in a Resource-Limited Setting

Nguyen Thi Thu Hong,[1] Nguyen To Anh,[1] Nguyen Thi Hoang Mai,[1] Ho Dang Trung Nghia,[1,2] Le Nguyen Truc Nhu,[1] Tran Tan Thanh,[1] Nguyen Hoan Phu,[1,3] Xutao Deng,[4,5] H. Rogier van Doorn,[1,6] Nguyen Van Vinh Chau,[7] Eric Delwart,[4,5] Guy Thwaites, [1,6] and Le Van Tan[1]

[1]Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam, [2]Pham Ngoc Thach University of Medicine, Ho Chi Minh City, Vietnam, [3]Department of Medicine, Vietnam National University, Ho Chi Minh City, Vietnam, [4]Vitalant Research Institute, San Francisco, California, USA, [5]Department of Laboratory Medicine, University of California, San Francisco, California, USA, [6]Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK, and [7]Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam

Background. Meningoencephalitis is a devastating disease worldwide. Current diagnosis fails to establish the cause in ≥50% of patients. Metagenomic next-generation sequencing (mNGS) has emerged as pan-pathogen assays for infectious diseases diagnosis, but few studies have been conducted in resource-limited settings.

Methods. We assessed the performance of mNGS in the cerebrospinal fluid (CSF) of 66 consecutively treated adults with meningoencephalitis in a tertiary referral hospital for infectious diseases in Vietnam, a resource-limited setting. All mNGS results were confirmed by viral-specific polymerase chain reaction (PCR). As a complementary analysis, 6 viral PCR-positive samples were analyzed using MinION-based metagenomics.

Results. Routine diagnosis could identify a virus in 15 (22.7%) patients, including herpes simplex virus (HSV; n = 7) and varicella zoster virus (VZV; n = 1) by PCR, and mumps virus (n = 4), dengue virus (DENV; n = 2), and Japanese encephalitis virus (JEV; n = 1) by serological diagnosis. mNGS detected HSV, VZV, and mumps virus in 5/7, 1/1, and 1/4 of the CSF positive by routine assays, respectively, but it detected DENV and JEV in none of the positive CSF. Additionally, mNGS detected enteroviruses in 7 patients of unknown cause. Metagenomic MinION-Nanopore sequencing could detect a virus in 5/6 PCR-positive CSF samples, including HSV in 1 CSF sample that was negative by mNGS, suggesting that the sensitivity of MinION is comparable with that of mNGS/PCR.

Conclusions. In a single assay, metagenomics could accurately detect a wide spectrum of neurotropic viruses in the CSF of meningoencephalitis patients. Further studies are needed to determine the value that real-time sequencing may contribute to the diagnosis and management of meningoencephalitis patients, especially in resource-limited settings where pathogen-specific assays are limited in number.

Keywords. metagenomics; next-generation sequencing; nanopore; MinION; meningoencephalitis.

Meningoencephalitis is a devastating clinical condition worldwide, but especially in tropical and resource-limited settings [1]. Although viruses are regarded as the most common causes of meningoencephalitis, the viruses responsible vary between geographic locations and are influenced by the emergence of pathogens such as Nipah virus, enterovirus A71, and Zika virus [2–4]. However, detecting many of these viruses is challenging, especially when most conventional diagnostic tests are pathogen specific (eg, polymerase chain reaction [PCR] for herpes simplex virus) and limited in number, especially in resource-limited settings. Even in well-equipped laboratories, a causative virus has only been established in <60% of patients [5–8].

Over the last decade, advanced sequencing technologies have emerged as a single pan-pathogen assay for the sensitive detection of known and unknown microorganisms, especially viruses, in cerebrospinal fluid (CSF) [6, 9, 10]. As part of our pathogen discovery, using a viral metagenomics-based approach, we previously identified a novel cyclovirus (CyCV-VN) in 4% of Vietnamese patients presenting with meningoencephaitis of unknown cause [11], although the pathogenic relevance of this novel circovirus species remains uncertain. From a diagnostic perspective, a recent prospective study in the United States compared the diagnostic performance of routine diagnostic tests with metagenomic next-generation sequencing (mNGS) and showed that mNGS detected a bacteria or virus in the CSF of 13 of 58 patients presenting with meningoencephalitis who were negative for or not assessed with routine diagnostic tests

Received 13 December 2019; editorial decision 4 February 2020; accepted 6 February 2020.

Correspondence: Le Van Tan, PhD, Oxford University Clinical Research Unit, 764 Vo Van Kiet, Ward 1, District 5, Ho Chi Minh City, Vietnam (tanlv@oucru.org).

Open Forum Infectious Diseases®

DOI: 10.1093/ofid/ofaa046

[6]. Otherwise, studies to date have been either case reports or retrospectively performed with small sample sizes [12], but few have been carried out in resource-limited settings like Vietnam. Such studies would have significant implications for both disease surveillance and patient management. Herein, we report the results of a study assessing the potential of metagenomics to detect a broad range of viruses in the CSF of consecutively treated adults with meningoencephalitis presenting to a tertiary referral hospital in southern Vietnam.

METHODS

Setting, Patient Enrollment, and Data Collection

The present study was conducted in a brain infection ward of the Hospital for Tropical Diseases (HTD) in Ho Chi Minh City, Vietnam, between January 2015 and September 2016. HTD is a tertiary referral hospital for patients, especially adults, with infectious diseases, including encephalitis, from the southern provinces of Vietnam with a population of >40 million.

One of the aims of the study was to improve diagnosis in patients with meningoencephalitis using metagenomic next-generation sequencing. We enrolled consecutive adult patients (≥18 years) with an indication for lumbar puncture admitted to the study site during the study period. Patients were excluded if pyogenic bacterial meningitis (cloudy or pus-like CSF) was suspected, lumbar puncture was contra-indicated, or no written informed consent was obtained was obtained from the patient or their relatives.

As per the study protocol, CSF samples were collected, alongside demographic and clinical data (including discharge outcome) and the results of routine diagnostic testing. After collection, all clinical specimens were stored at –80°C for subsequent analyses, including assessment of mNGS performance against that of routine diagnostic assays. Here we focused our analysis on patients with meningoencephalitis regardless of the results of routine diagnosis. Additionally, as negative controls, 1 CSF from a patient presenting with cerebral hemorrhage and 1 from a patient with laboratory-confirmed anti-N-methyl-D-aspartate receptor [13] were also included.

Routine Diagnosis

As part of routine care at HTD, CSF specimens of patients presenting with brain infections were cultured and/or examined by microscopy for detection of bacterial/fungal/*Mycobacterium tuberculosis* infection with the use of standard methods when appropriate (Supplementary Table 1). Herpes simplex virus (HSV) PCR was carried out in patients presenting with clinically suspected meningoencephalitis. Varicella zoster virus (VZV) PCR, serological testing for IgM against dengue virus (DENV), Japanese encephalitis virus (JEV), or MuV was performed if clinically indicated and testing for other pathogens (HSV) was negative [8].

Illumina MiSeq/MinION–Based Viral Metagenomics

Sample Pretreatments and Nucleic Acid Isolation

To allow for the detection of both RNA and DNA viruses, each CSF sample was subjected to 2 different metagenomic approaches, namely RNA virus and viral DNA virus workflows (Figure 1). For the former, 200 µL of CSF was first pretreated with 2 U/µL of turbo DNase (Ambion, Life Technology, Carlsbad, CA, USA) and 0.4 U/µL RNase 1 (Ambion) at 37°C for 30 minutes by DNase and RNase, followed by nucleic acid (NA) isolation using the QIAamp viral RNA kit (QIAgen GmbH, Hilden, Germany). For the latter, viral DNA was directly isolated from 200 µL of CSF samples without the nuclease treatment step using the DNeasy blood and tissue kit (QIAgen GmbH). Finally, viral RNA/DNA of both workflows was recovered in 50 µL of elution buffer.

Double-Stranded DNA Synthesis and Random Amplification of Extracted Viral RNA

Double-stranded DNA was synthesized from isolated viral RNA using a set of 96 nonribosomal random primer, followed by PCR amplification to enrich for viral RNA before sequencing as previously described [14–16]. In brief, 10 µL of extracted viral RNA was converted into double-stranded DNA (dsDNA) using FR26RV-Endoh primers [16], Super Script III enzyme (Invitrogen, Carlsbad, CA, USA), RNase OUT (Invitrogen), exo-Klenow fragment (Ambion), and Ribonuclease H (Ambion). Subsequently, the synthesized dsDNA was randomly amplified using FR20RV primer (5'-GCCGGAGCTCTGCAGATATC-3'). The obtained random PCR product was then purified with use of Agencourt AMPure XP beads (Beckman coulter) and quantified using the Qubit dsDNA HS kit (Invitrogen).

Next-Generation Sequencing

One ng of the purified random PCR product of the RNA virus workflow and isolated viral DNA of the DNA virus workflow was subjected to the library preparation step using the Nextera XT sample preparation kit (Illumina, San Diego, CA, USA), following the manufacturer's instructions. Samples were multiplexed using the combinatorial indexing strategy (ie, only 1 index might be shared between samples). The resulting libraries of both workflows were separately sequenced using MiSeq reagent kits, version 3 (600 cycles; Illumina), in a MiSeq platform (Illumina), following the manufacturer's instructions. All the experiments were performed in molecular diagnostic facilities that consist of 3 physically separated laboratories for reagent preparation, extraction, and library preparation and sequencing. These were used a unidirectional workflow.

MinION Library Preparation and Sequencing

A subset of 6 CSF samples in which a virus was detected by PCR and/or mNGS was selected for a complementary analysis using MinION sequencer (Oxford Nanopore Technologies). MinION libraries were prepared using either extracted DNA or random

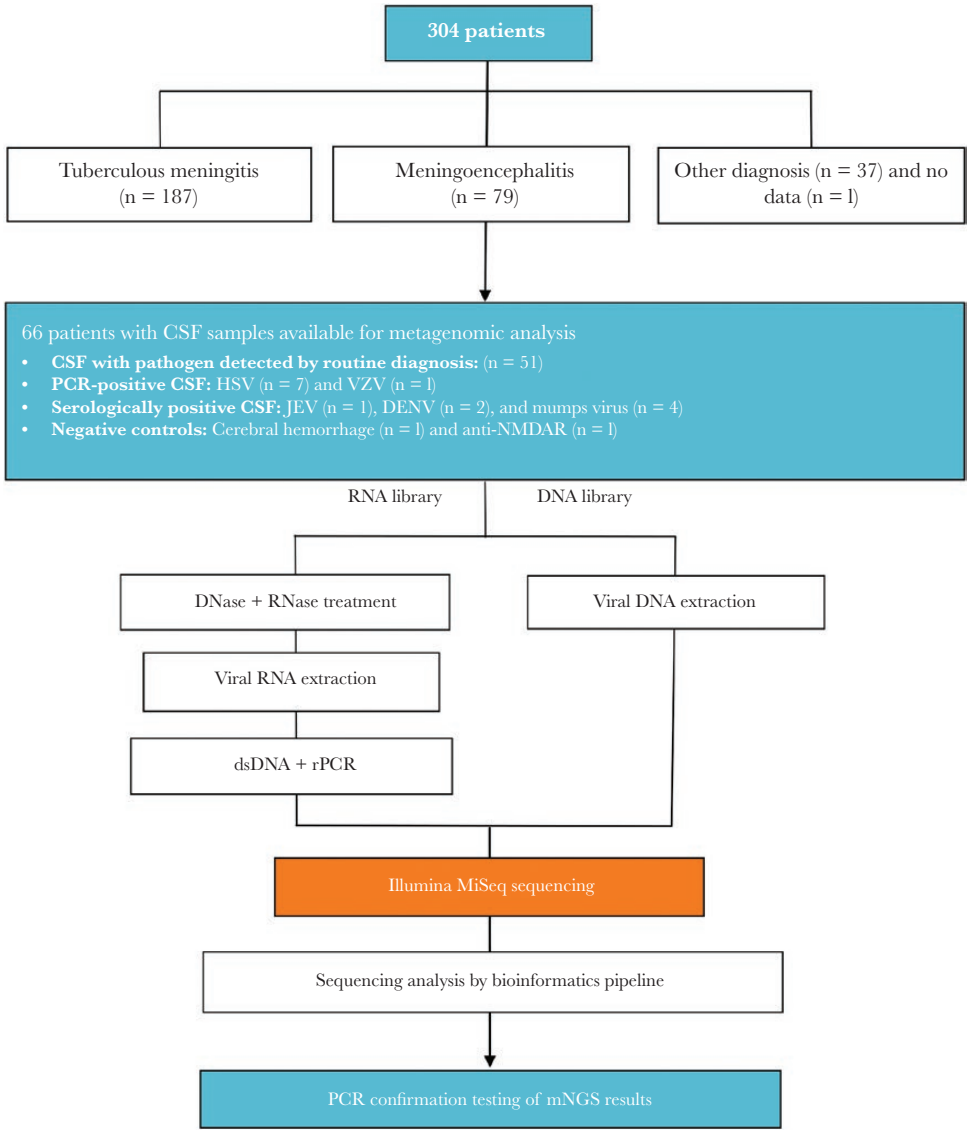


Figure 1. Flowchart illustrating an overview about the DNA and RNA virus workflows. Abbreviations: CSF, cerebrospinal fluid; DENV, dengue virus; ds, double-stranded; JEV, Japanese encephalitis virus; mNGS, metagenomic next-generation sequencing; PCR, polymerase chain reaction.

amplified products synthesized as described above using the 1D Native Barcoding Genomic DNA kit (ONT, Oxford, UK), following the manufacturer's protocol. The 6 CSF samples and a nontemplate control (each was assigned to unique barcodes) were sequenced in 1 single run using R9.4 flow cells (ONT). Base-calling of MinION reads was performed using MinKNOW (ONT), followed by demultiplexing of the obtained reads using Porechop (https://github.com/rrwick/Porechop).

Sequence Analysis of the Obtained Metagenomic Reads

The mNGS data generated by the Illumina MiSeq platform were analyzed using an in-house viral metagenomic pipeline running on a 36-node Linux cluster available through Vitalant Research Institute, San Francisco, to identify the presence of viral sequences in the tested specimens, as previously described [17, 18]. In brief, after filtering out duplicate reads and reads belonging to human and bacterial genomes, and with adaptors and low-quality reads trimmed, the remaining reads were de novo assembled. The resulting contigs and singlet reads were then aligned against a customized viral proteome database extracted from the NCBI's RefSeq and NR databases using a Basic Local Alignment Search Tool (BLAST)–based approach. Next, the candidate viral reads were aligned against a nonredundant nonvirus protein database to remove any false-positive reads (ie, reads with expected [E] values higher than those in viral protein databases) using DIAMOND [19]. Any viral-like sequence with an E value of $\leq 10^{-5}$ was considered a significant hit and was then manually checked by BLASTX to further exclude false-positive hits. Finally, a reference-based mapping approach was employed to assess the level of identity and genome coverage of the corresponding viruses.

Analysis of MinION reads was carried out using Taxonomer [20], a publicly available metagenomics pipeline, which incorporates an interactive results visualization function.

PCR Confirmation of Viral Hits Detected by Metagenomics and Expanded PCR Testing

Because of the uncertainty in the diagnostic performance of mNGS and the focus of the present study, we performed specific PCRs to confirm mNGS hits matched with the genomes of neurotropic viruses. The PCR experiments were either carried out on leftover extracted RNA/DNA after the mNGS library preparation experiments or on newly extracted nucleic acids (NA). An mNGS result was only considered positive if it was subsequently confirmed by a corresponding viral PCR analysis of the original NA materials derived from corresponding individual samples. All PCR primers and probes used were derived from previous publications [21–23], including a real-time reverse transcription PCR (RT-PCR) for generic detection of enteroviruses.

Because of the focus of the present study, viruses of unknown neurotropic property and well-known contaminants of the mNGS data set were not pursued further by subsequent PCR analysis.

Unless specified above, all the laboratory experiments and bioinformatics analyses were carried out at the Oxford University Clinical Research Unit in Ho Chi Minh City, Vietnam.

GenBank Accession Numbers

Metagenomics data were deposited at NCBI (GenBank) under SRA accession number PRJNA58865.

Ethics

This clinical study received approvals from the Institutional Review Board of the HTD and the Oxford Tropical Research Ethics Committee of the University of Oxford. Written informed consent was obtained from each study participant or relative (if the patient was unconscious).

RESULTS

Baseline Characteristics of the Patients Included for mNGS

During the study period, a total of 304 patients were enrolled in the clinical study, including patients with tuberculous meningitis (n = 187), meningoencephalitis (n = 79), another
berculous meningitis diagnostic arm have been published elsewhere [24]. Of the 79 patients with a discharge diagnosis of meningoencephalitis, 66 (84%) had CSF samples available for metagenomic analysis (Figure 1). These patients were the focus of the present study regardless of the results of routine diagnosis.

The baseline characteristics of the 66 patients included in the study are presented in Table 1. HIV testing was carried out in 24 patients; none were positive. Males were predominant. On admission, 35% of the patients were comatose (Glasgow Coma Score < 13). Routine diagnostic tests identified a virus in 15/66 (22.7%) patients (Figure 2; Supplementary Table 2), with HSV being the most common cause (n = 7), followed by MuV (n = 4), DENV (n = 2), JEV (n = 1), and VZV (n = 1) (Figure 2). One patient died, and almost all (n = 58) had some neurological deficit at discharge from the hospital (Table 1).

An Overview of mNGS

The 68 included CSF samples (including 2 negative controls) were separately sequenced using both DNA and RNA virus workflows in a blinded fashion. Subsequently, a total of 62 565 802 and 49 233 869 reads were obtained from the DNA and RNA libraries, respectively (Supplementary Table 3). Sequences related to 29 viral species were detected, with 23 found in the RNA and 7 found in the DNA library (Figures 2 and 3). The detected viruses included viruses known to cause CNS infections and those with unknown neurotropic properties (Torque teno virus [n = 14] and herpes virus 8 [n = 4]). Additionally, previously reported common contaminants of the mNGS data set were also found [25, 26], almost exclusively in the RNA virus library (Figure 3).

Detection of Viruses in CSF Samples That Were Positive by Routine Diagnosis

Of the 15 CSF samples positive either by PCR or serological testing as part of routine care, mNGS was able to detect a viral pathogen in 5/7 HSV-, 1/1 VZV-, 1/4 MuV-, 0/2 DENV-, and 0/1 JEV-positive samples (Figure 2). None of the HSV and VZV sequences were found in the library of the RNA virus workflow (Table 2).

Detection of sequences related to human pathogenic viruses in CSF that were negative by routine diagnosis, and results of PCR assessment of mNGS results

Of the 51 CSF samples that were negative by routine diagnosis, sequences related to neurotropic viruses were found in 24 (48%) samples by mNGS (Table 2). The detected viruses included enteroviruses (EVs; n = 23) and rotavirus (n = 1). Additionally, of the 2 CSF samples from non-CNS-affected patients, 1 had 4 sequences related to enterovirus detected by mNGS.

After PCR confirmation testing of CSF samples in which a viral hit was detected by mNGS, the rotavirus case and the negative control CSF, in which EV-related sequences were detected, became negative (Table 2). The number of EV-positive CSF samples was reduced from 23 to 7, with more enteroviral sequences being recorded in the PCR-confirmed group than in the unconfirmed group (Table 2). Of these, 3 had genome coverage of 61%, 78%, and 90%, including 1 echovirus 6 and 2 echovirus 30. Notably, the majority (12/16, 75%) of EV PCR-negative samples had EV reads identical to those obtained from samples with a high abundance of EV sequences (including samples #12 and #14), with which they shared an index (Supplementary Table 4), suggesting the potential of barcode bleedthrough during the sequencing procedure.

Table 1. Baseline Characteristics of the Study Patients and Patients Infected With HSV/EVs/Mumps Virus

	Total (n = 66)[a]	HSV (n = 7)[b]	EVs (n = 7)[c]	Mumps Virus (n = 5)[d]
Demographics				
Gender (male), No. (%)	39 (59)	4 (57)	5/7 (71)	5 (100)
Age, y	35 (15–84)	45 (25–53)	32 (22–57)	39 (32–61)
Illness day on admission, d	5 (1–30)	5 (2–14)	3.5 (2–6)	3 (2–5)
Duration of hospital stay, d	5 (1–76)	5 (3–67)	2 (1–4)	4 (3–35)
HIV status, No. (%)				
Positive	0	0	0	0
Negative	24 (36)	1 (14)	4 (57)	1 (20)
Unknown	42 (64)	6 (86)	3 (43)	4 (80)
Clinical signs and symptoms, No. (%)				
Fever	58 (88)	7/7 (100)	6 /7 (86)	5 (100)
Headache	58 (88)	7/7 (100)	6 /7 (86)	5 (100)
Irritability	15 (23)	1/7 (14)	1/7 (14)	0
Lethargy	18 (28)	3/6 (50)	1/7 (14)	0
Vomiting	34 (52)	4/6 (67)	5/7 (71)	3 (60)
Seizures	23 (36)	2/6 (33)	0/7	2 (40)
Conscious	46 (70)	6/7 (86)	1/7 (14)	2 (40)
Skin rash	6 (9)	0/7	0/7	0
Hemiplegia	5 (8)	2/7 (29)	0/7	0
Paraplegia	1 (2)	0/7	1/7 (14)	0
Tetraplegia	1 (2)	0/6	0/7	0
Neck stiffness	45 (68)	6/7 (86)	5/7 (71)	3 (60)
Glasgow coma score of ≤8	7 (11)	3/7 (43)	0/7	1 (20)
Glasgow coma score of 9–12	16 (24)	2/7 (29)	1/7 (14)	1 (20)
Glasgow coma score of 13–15	43 (65)	2/7 (29)	6 /7 (86)	3 (60)
CSF cells and biochemistry				
White cells, cells/μL	101 (0–4183)	708 (38–1571)	503 (20–961)	683 (27–2146)
Neutrophils, No. (%)	13 (0–96)	9 (2–61)	24 (0–47)	18 (3–23)
Lymphocytes, No. (%)	86.5 (1–100)	91 (39–98)	76 (53–99.9)	82 (77–97)
Protein, g/L	0.7 (0.2–8.9)	1.36 (0.75–2.17)	0.71 (0.47–1.18)	0.67 (0.45–2.42)
CSF/blood glucose ratio	0.61 (0.34–1.04)	0.55 (0.47–0.61)	0.71 (0.59–0.85)	0.52 (0.49–0.81)
Lactate, mmol/L	2.65 (1.4–14.03)	3.52 (2.02–4.83)	2.5 (1.9–3.8)	2.9 (1.9–4.3)
Antiviral treatment, No. (%)				
Oral acyclovir	2 (3)	NA	NA	NA
Intravenous acyclovir	8 (13)	6/6 (100)	NA	NA
Oral valacyclovir	44 (72)	NA	NA	1 (20)
Modified Rankin Scale at discharge,[e] No. (%)				
0	8 (13)	1/7 (14)	1/7 (14)	1 (20)
1	12 (19)	0	1/7 (14)	3 (60)
2	10 (15)	0	4/7 (58)	0
3	25 (39)	3/7 (43)	1/7 (14)	1 (20)
4	4 (6)	0	0	0
5	4 (6)	3/7 (43)	0	0
6	1 (2)	0	0	0

Continuous variables are presented as median (range).

[a]Denominators may vary slightly.

[b]Diagnosed by current standard tests for routine diagnosis.

[c]Diagnosed by mNGS, followed by PCR confirmatory testing.

[d]Diagnosed by current standard tests, expanded PCR testing, and mNGS combined.

[e]0: Full recovery with no symptoms; 1: No significant disability; 2: Slight disability; 3: Moderate disability; 4: Moderately severe disability; 5: Severe disability; and 6: Dead.

Results of Expanded PCR Testing and Sensitivity Assessment of mNGS Using PCRs as Reference Assays

Because PCR testing for viruses (EVs and MuV) was not performed as part of routine diagnosis, to further assess the prevalence of these viruses in the study patients, we expanded PCR testing to CSF samples that were negative by mNGS analysis. Subsequently, only MuV was detected by PCR in 4 CSF samples, including 1 positive by both serological and mNGS methods (real-time PCR cycle threshold [Ct] values: 35), 2 positive by serological testing as part of standard care (Ct values: 36 and 40), and 1 negative by mNGS (Ct value: 40). Serological testing for MuV in

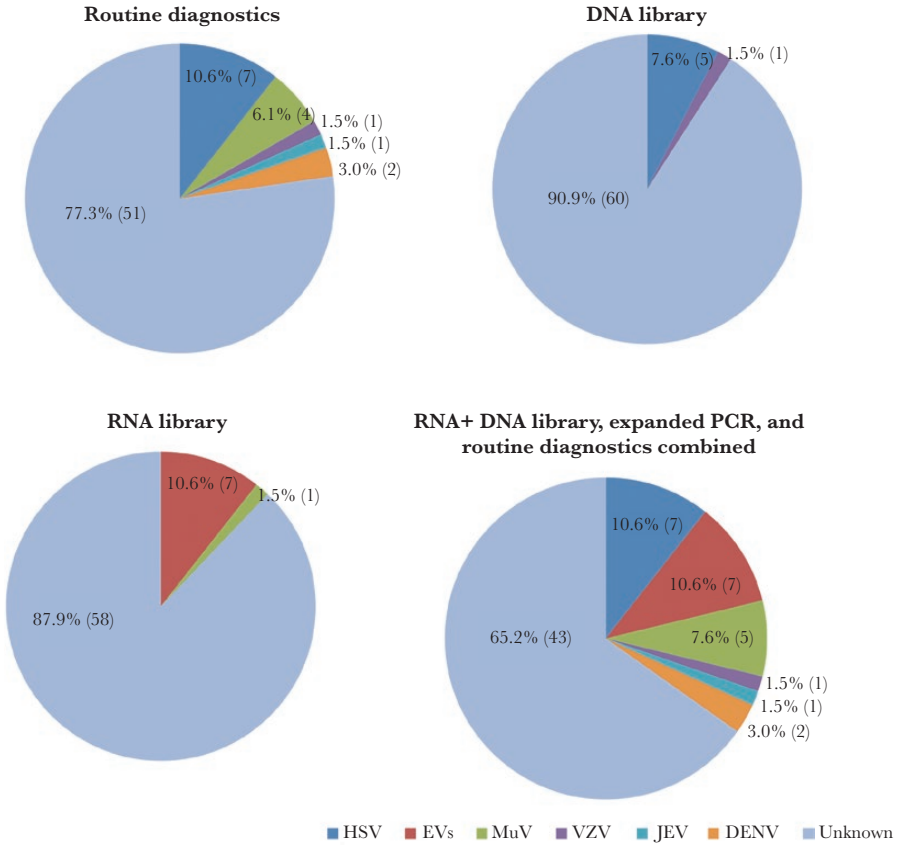


Figure 2. Results of metagenomic investigations using DNA/RNA workflows and routine diagnostics as well as expanded polymerase chain reaction testing. Abbreviations: DENV, dengue virus; EV, enterovirus; HSV, herpes simplex virus; JEV, Japanese encephalitis virus; MuV, mumps virus; PCR, polymerase chain reaction; VZV, varicella zoster virus.

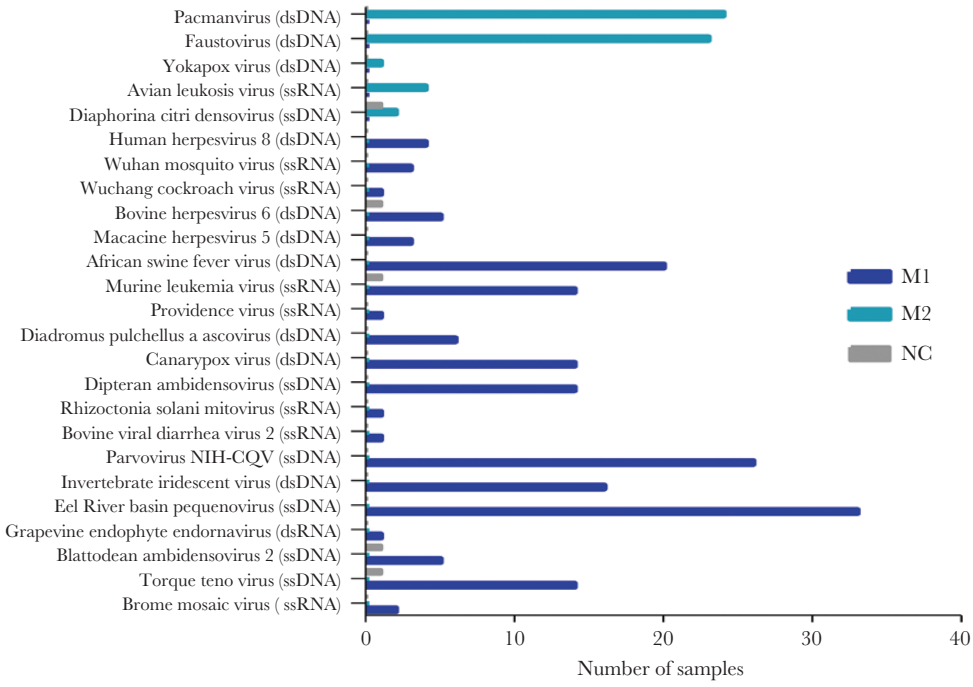


Figure 3. Bar chart showing the frequency of common contaminants and viruses of unknown neurotropic property (human herpes virus 8 and Torque teno virus) found in cerebrospinal fluid (CSF) samples by both DNA and RNA workflow and viruses in negative control CSF. Abbreviations: ds, double-stranded; ss, single-stranded.

Table 2. Results of Viral PCR and Metagenomic Analysis

CSF No.	Virus	Real-time PCR Ct Value	Detected by PCR as Part of Routine Care (Y/N)	Total Metagenomic Reads	No. of Unique Viral Reads	(%) of Viral Reads[a]	mNGS Library
1	HSV	25.01	Y	326 396	49	0.015	DNA
2	HSV	28.01	Y	588 504	184	0.031	DNA
3	HSV	30.36	Y	996 348	6	0.001	DNA
4	HSV	23.77	Y	1 145 710	243	0.021	DNA
5	HSV	28.71	Y	346 166	11	0.003	DNA
6	HSV	Unavailable	Y	1 345 954	0	0.000	NA
7	HSV	31	Y	891 566	0	0.000	NA
8	VZV	22.7	Y	1 335 288	152	0.011	DNA
9	Mumps	35.2	Y	975 714	6	0.001	RNA
10	Enterovirus	33.36	ND	539 752	21	0.004	RNA
11	Enterovirus	34.25	ND	635 310	38	0.006	RNA
12	Enterovirus	34.79	ND	765 564	10152	1.326	RNA
13	Enterovirus	34.78	ND	732 634	89	0.012	RNA
14	Enterovirus	31.23	ND	988 668	2415	0.244	RNA
15	Enterovirus	32.3	ND	594 964	100	0.017	RNA
16	Enterovirus	35.65	ND	543 912	21	0.004	RNA
17	Enterovirus	Negative	ND	579 486	2	0.000	RNA
18	Enterovirus	Negative	ND	571 902	2	0.000	RNA
19	Enterovirus	Negative	ND	720 042	4	0.001	RNA
20	Enterovirus	Negative	ND	511 608	1	0.000	RNA
21	Enterovirus	Negative	ND	818 654	2	0.000	RNA
22	Enterovirus	Negative	ND	513 428	5	0.001	RNA
23	Enterovirus	Negative	ND	1 197 290	13	0.001	RNA
24	Enterovirus	Negative	ND	923 908	4	0.000	RNA
25	Enterovirus	Negative	ND	993 918	1	0.000	RNA
26	Enterovirus	Negative	ND	1 302 784	20	0.002	RNA
27	Enterovirus	Negative	ND	1 628 722	7	0.000	RNA
28	Enterovirus	Negative	ND	1 181 716	24	0.002	RNA
29	Enterovirus	Negative	ND	926 462	22	0.002	RNA
30	Enterovirus	Negative	ND	938 524	20	0.002	RNA
31	Enterovirus	Negative	ND	1 028 194	12	0.001	RNA
32	Enterovirus	Negative	ND	1 239 458	4	0.000	RNA
33	Rotavirus	Negative	ND	1 176 486	24	0.002	RNA

[a]Denominators are the total reads of the corresponding samples.

Abbreviations: CSF, cerebrospinal fluid; mNGS, metagenomic next-generation sequencing; ND, not done; PCR, polymerase chain reaction.

this patient was not done as part of routine care (Supplementary Table 5). Thus a combination of serology and molecular assays (PCR and mNGS) increased the diagnostic yield from 22.7% (15/66) to 34.8% (23/66) (Figure 2).

mNGS identified a viral pathogen in 14/19 CSF samples that were positive by PCR analysis (including routine diagnosis and expanded testing). Additionally, mNGS detected reads related to EVs in 16/47 CSF samples that were negative by subsequent PCR analysis. Using PCRs as reference assays, the sensitivity and specificity of mNGS were 74% (14/19) and 66% (31/47), respectively. Of the PCR-positive samples, there was no difference in the leukocyte counts between the mNGS-negative and -positive groups (median [range], 331 [27–2146] vs 356 [22–4183]; $P = .82$).

Rapid Detection of Encephalitis in CSF by MinION Nanopore Sequencing

A MinION Nanopore–based metagenomic approach detected HSV (n = 2), VZV (n = 1), and EV (n = 2) in 5/6 CSF samples that were PCR positive for these viruses (Figure 4A). Of these 5 MinION-positive samples, 1 HSV sample was negative, and the other 4 were positive by MiSeq-based mNGS workflows (Figure 4A). Notably, after 2 hours of the sequencing run, reads assigned to corresponding viral species found in CSF by PCR were obtained in 4/5 MinION-positive samples. MinION, however, failed to detect MuV in a CSF sample that was positive by both PCR (Ct value = 36) and MiSeq workflow (Figure 4).

DISCUSSION

We report the results of an investigation assessing the utility of next/third-generation sequencing–based metagenomics as a hypothesis-free approach for detection of viral etiology in the CSF of 66 consecutively treated patients with meningoencephalitis. The patients were admitted to a tertiary referral hospital in Ho Chi Minh City, Vietnam, and the majority (51%) had moderate/severe disability at discharge. The results showed that in a

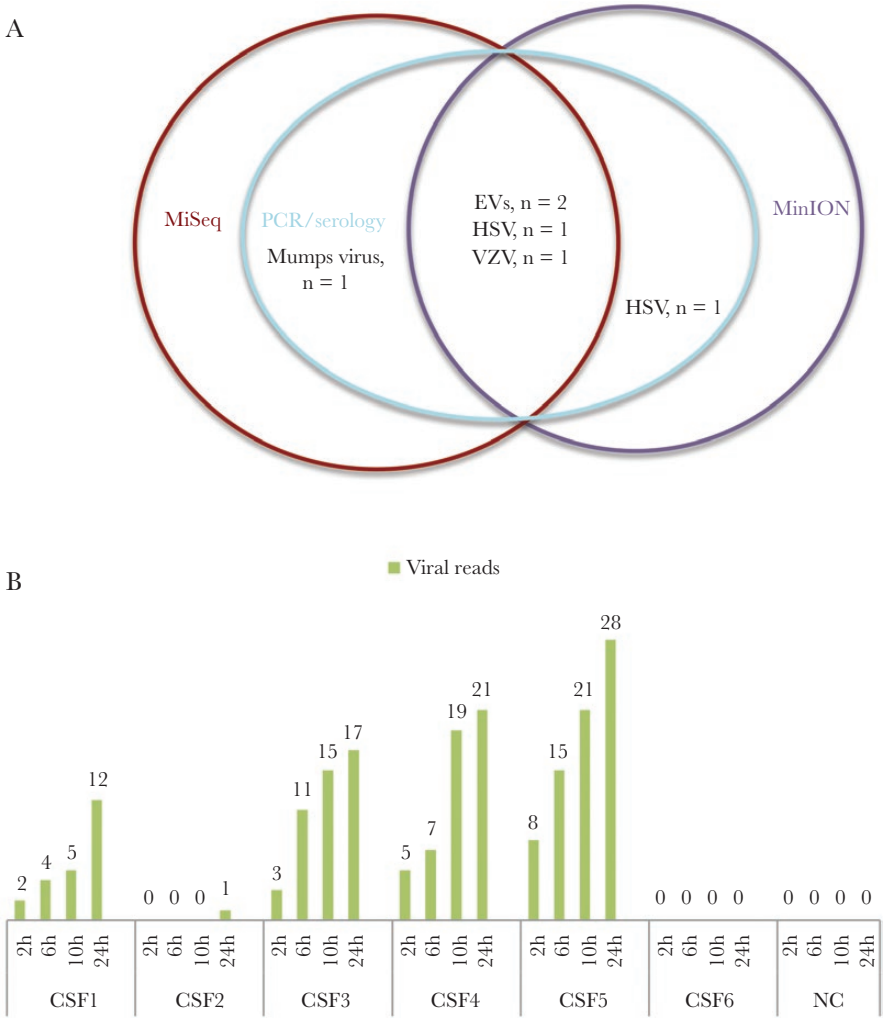


Figure 4. Results of MinION Nanopore–based metagenomics. A, Venn diagram showing the agreement between metagenomic approaches and polymerase chain reaction (PCR). B, Cumulative number of MinION reads assigned to corresponding viral species found in cerebrospinal fluid (CSF) by PCR at different time points. CSF1 and CSF2: herpes simplex virus (HSV) positive; CSF3: varicella zoster virus positive; CSF4 and CSF5: enterovirus positive; CSF6: mumps virus positive. All the 12 reads obtained from CSF1 were assigned to HSV1, and the single reads obtained from CSF2 were assigned to HSV2. Abbreviations: EV, enterovirus; HSV, herpes simplex virus; PCR, polymerase chain reaction; VZV, varicella zoster virus.

single test metagenomics could accurately detect nucleic acids of a wide range of neurotropic viruses in the CSF of 66 participants, whose diagnoses were only established by extensive PCR testing targeted at a broad range of pathogens. Notably, of these 66 patients, 7 (11%) EV-infected patients were initially left undiagnosed at hospital discharge because physicians did not consider EV diagnosis as part of routine care. EV infection should therefore be considered as an important differential diagnosis in adults presenting with meningoencephalitis [27] and should be excluded (eg, by PCR testing) before mNGS analysis.

Although antivirals are currently not available for most encephalitis-causing viruses, rapid and accurate detection of viral etiology in patient samples remain essential to inform clinical management, such as avoiding unnecessary antibiotic prescription, and public health policy-makers. Thus, testing for a wide spectrum of pathogens is essential to maximize the diagnostic yield in patients presenting with meningoencephalitis. Under these circumstances, a single pan-pathogen assay such as mNGS is a useful approach, given the limited amount of CSF samples and resources available for microbial investigation, especially in low- and middle-income countries like Vietnam. However, the failure of mNGS to detect nucleic acids of JEV and DENV in serologically positive CSF samples emphasizes that testing for pathogen-specific antibodies remains an important diagnostic pathway in patients presenting with meningoencephalitis, as viral nucleic acids of some viruses (eg, flaviviruses) may not be present in the collected CSF.

The sensitivity of our mNGS workflows is comparable with that of recent mNGS studies [6, 9]. Low viral load may be a factor in the failure of mNGS to detect HSV and MuV in CSF samples with real-time PCR Ct values of 31 for HSV and 36, 40, and 40 for MuV. Because viral reads only accounted for a small

proportion of total mNGS reads, increasing the sequencing depth per sample would likely increase the sensitivity of mNGS. However, this increases the sequencing costs.

Currently, there are no established robust criteria that can reliably define a true mNGS positive without the requirement of confirmatory testing. Criteria such as the presence of at least 3 reads mapped to 3 different genomic regions of a virus genome or the absence of viral reads in negative controls have recently been proposed [6, 10, 12]. Such approaches are hindered by the well-known cross-talk contamination phenomenon, occurring as part of the mNGS procedure [10], which, however, can be dramatically reduced through the use of the dual barcoding strategy, which was recently developed [28]. Because we did not employ the dual barcoding strategy, cross-talk contamination may explain the obtained specificity of 66%, which is lower than the reported data from a previous study [9]. Alternatively, the low specificity could be attributed to the degradation of stored viral RNA and/or the low abundance of viral RNA in the tested samples, leading to the failure of EV PCR to replicate some of the mNGS findings. Retrospectively, the specificity of mNGS would have increased to 83% if a threshold of ≥6 reads was considered positive (Table 2), suggesting a correlation between the number of mNGS reads and PCR confirmatory results. Collectively, the specificity of the mNGS-based diagnostic approach could potentially be improved through the use of a proper barcoding strategy and/or criteria such as those based on the number of unique viral reads obtained from a sample under investigation, which merits further research.

Recently, the single-molecule real-time sequencing developed by Oxford Nanopore Technologies has emerged as a promising tool for clinical settings because of its short turnaround time. As such, it could potentially overcome the current limitation of the long turnaround time posed by other NGS platforms. However, scarce information exists for the application of Oxford Nanopore Technologies as a hypothesis-free approach to detect viral agents in clinical samples [10, 29, 30]. The results of our complementary analysis demonstrate that MinION-based metagenomics could accurately detect viral pathogens in CSF samples within 2 hours after the sequencing run, whereas the current Illumina MiSeq–based metagenomic approach takes around 48–56 hours to complete. Collectively, the data suggest that the sensitivity of MinION is comparable with that of mNGS/PCR, and thus point to the utility potential of MinION sequencing for rapid diagnosis of meningoencephalitis, which merits further research.

Similar to previous reports [25, 26], numerous common contaminants of the mNGS data set (eg, parvovirus, densovirus) were found in both the DNA and RNA virus libraries in our study. Although it is likely that those contaminants were derived from laboratory reagents (eg, extraction kits) [25], their potential impacts on the performance of mNGS, especially in terms of sensitivity and specificity, remain unknown.

The strengths of our study include that it was conducted on consecutive cases, minimizing selection bias. CSF samples were analyzed individually, and mNGS hits were reconfirmed by specific PCR, allowing for back-to-back comparison between mNGS and viral PCR. However, our study has some limitations. First, it was conducted on stored CSF samples. Second, we only focused on viruses, while meningoencephalitis can be caused by nonviral agents such as intracellular bacteria (rickettsiae) [31]. Third, we did not test other clinical samples. Of note, JEV has recently been detected in the urine of patients presenting with meningoencephalitis [32, 33]. Last but not least, the inclusion of nontemplate controls in addition to the 2 noninfectious CSF samples would have better captured the spectrum of contaminations of the mNGS procedure.

To summarize, we report pioneering data on the performance of metagenomic next/third-generation sequencing on the CSF of meningoencephalitis patients in Vietnam, a resource-limited setting. The results shows that in a single assay, metagenomics was able to detect a wide spectrum of neurotropic viruses in CSF samples of meningoencephalitis patients, and thus it could potentially replace conventional nucleic acid–based diagnostic assays such as PCR. Further studies are needed to determine the clinical implications of real-time sequencing in the diagnosis and management of meningoencephalitis patients, especially in resource-limited settings, where pathogen-specific assays are limited in number.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments

We thank Le Kim Thanh for her logistic support. We are indebted to the patients for their participations in this study.

Financial support. This study was funded by the Wellcome Trust of Great Britain (106680/B/14/Z and 204904/Z/16/Z).

Potential conflicts of interest. All authors: No reported conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Roth GA, Abate D, Abate KH, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet **2018**; 392(10159):1736–88.
2. Musso D, Ko AI, Baud D. Zika virus infection - after the pandemic. N Engl J Med **2019**; 381:1444–57.
3. Paton NI, Leo YS, Zaki SR, et al. Outbreak of Nipah-virus infection among abattoir workers in Singapore. Lancet **1999**; 354:1253–6.
4. Ooi MH, Wong SC, Lewthwaite P, et al. Clinical features, diagnosis, and management of enterovirus 71. Lancet Neurol **2010**; 9:1097–105.
5. Granerod J, Ambrose HE, Davies NW, et al; UK Health Protection Agency (HPA) Aetiology of Encephalitis Study Group. Causes of encephalitis and differences in their clinical presentations in England: a multicentre, population-based prospective study. Lancet Infect Dis **2010**; 10:835–44.

6. Wilson MR, Sample HA, Zorn KC, et al. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. N Engl J Med **2019**; 380:2327–40.
7. Glaser CA, Gilliam S, Schnurr D, et al. In search of encephalitis etiologies: diagnostic challenges in the California Encephalitis Project, 1998–2000. Clin Infect Dis **2003**; 36(6):731–42.
8. Tan le V, Thai le H, Phu NH, et al. Viral aetiology of central nervous system infections in adults admitted to a tertiary referral hospital in southern Vietnam over 12 years. PLoS Negl Trop Dis **2014**; 8:e3127.
9. Miller S, Naccache SN, Samayoa E, et al. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. Genome Res **2019**; 29:831–42.
10. Chiu CY, Miller SA. Clinical metagenomics. Nat Rev Genet **2019**; 20:341–55.
11. Tan le V, van Doorn HR, Nghia HD, et al. Identification of a new cyclovirus in cerebrospinal fluid of patients with acute central nervous system infections. mBio **2013**; 4:e00231–13.
12. Brown JR, Bharucha T, Breuer J. Encephalitis diagnosis using metagenomics: application of next generation sequencing for undiagnosed cases. J Infect **2018**; 76:225–40.
13. Nguyen Thi Hoang M, Nguyen Hoan P, Le Van T, et al. First reported cases of anti-NMDA receptor encephalitis in Vietnamese adolescents and adults. J Neurol Sci **2017**; 373:250–3.
14. Endoh D, Mizutani T, Kirisawa R, et al. Species-independent detection of RNA virus by representational difference analysis using non-ribosomal hexanucleotides for reverse transcription. Nucleic Acids Res **2005**; 33:e65.
15. Anh NT, Hong NTT, Nhu LNT, et al. Viruses in Vietnamese patients presenting with community-acquired sepsis of unknown cause. J Clin Microbiol **2019**; 57(9):1–13.
16. Nguyen AT, Tran TT, Hoang VM, et al. Development and evaluation of a non-ribosomal random PCR and next-generation sequencing based assay for detection and sequencing of hand, foot and mouth disease pathogens. Virol J **2016**; 13:125.
17. Deng X, Naccache SN, Ng T, et al. An ensemble strategy that significantly improves de novo assembly of microbial genomes from metagenomic next-generation sequencing data. Nucleic Acids Res **2015**; 43:e46.
18. Li L, Deng X, Mee ET, et al. Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. J Virol Methods **2015**; 213:139–46.
19. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods **2015**; 12:59–60.
20. Flygare S, Simmon K, Miller C, et al. Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling. Genome Biol **2016**; 17:1–18.
21. Thanh TT, Anh NT, Tham NT, et al. Validation and utilization of an internally controlled multiplex real-time RT-PCR assay for simultaneous detection of enteroviruses and enterovirus A71 associated with hand foot and mouth disease. Virol J **2015**; 12:1–9.
22. Uchida K, Shinohara M, Shimada S, et al. Rapid and sensitive detection of mumps virus RNA directly from clinical samples by real-time PCR. J Med Virol **2005**; 75:470–4.
23. Dung TT, Phat VV, Nga TV, et al. The validation and utility of a quantitative one-step multiplex RT real-time PCR targeting rotavirus A and norovirus. J Virol Methods **2013**; 187:138–43.
24. Heemskerk AD, Donovan J, Thu DDA, et al. Improving the microbiological diagnosis of tuberculous meningitis: a prospective, international, multicentre comparison of conventional and modified Ziehl-Neelsen stain, GeneXpert, and culture of cerebrospinal fluid. J Infect **2018**; 77:509–15.
25. Asplund M, Kjartansdóttir KR, Mollerup S, et al. Contaminating viral sequences in high-throughput sequencing viromics: a linkage study of 700 sequencing libraries. Clin Microbiol Infect **2019**; 25:1277–85.
26. Xu B, Zhi N, Hu G, et al. Hybrid DNA virus in Chinese patients with seronegative hepatitis discovered by deep sequencing. Proc Natl Acad Sci U S A **2013**; 110:10264–9.
27. Glaser CA, Honarmand S, Anderson LJ, et al. Beyond viruses: clinical profiles and etiologies associated with encephalitis. Clin Infect Dis **2006**; 43:1565–77.
28. Wilson MR, Fedewa G, Stenglein MD, et al. Multiplexed metagenomic deep sequencing to analyze the composition of high-priority pathogen reagents. mSystems **2016**; 1(4):1–9.
29. Greninger AL, Naccache SN, Federman S, et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med **2015**; 7:1–13.
30. Mongan AE, Tuda JSB, Runtuwene LR. Portable sequencer in the fight against infectious disease. J Hum Genet **2020**; 65:35–40.
31. Sekeyová Z, Danchenko M, Filipčík P, Fournier PE. Rickettsial infections of the central nervous system. PLoS Negl Trop Dis **2019**; 13:e0007469.
32. Huang GKL, Tio SY, Caly L, et al. Prolonged detection of Japanese encephalitis virus in urine and whole blood in a returned short-term traveler. Open Forum Infect Dis **2017**; 4(X):XXX–XX.
33. Mai NTH, Phu NH, Nhu LNT, et al. Central nervous system infection diagnosis by next-generation sequencing: a glimpse into the future? Open Forum Infect Dis **2017**; 4(X):XXX–XX.